

**“Generalized Entropies and Legendre Duality”**

**22/04/2012**

**Name of Principal Investigators:** Keiko Uohashi

- e-mail address : uohashi@tjcc.tohoku-gakuin.ac.jp
- Institution : Department of Mechanical Engineering & Intelligent Systems,  
Faculty of Engineering, Tohoku Gakuin University
- Mailing Address : 1-13-1 Chuo, Tagajo, Miyagi 985-8537, Japan
- Phone : +81-22-368-7284
- Fax : +81-22-368-7070

Period of Performance: 22/04/2010 – 22/04/2011

**Abstract:** Making use of conformally flattened structure of alpha-geometry, we have shown that the simple and computationally efficient algorithm can be derived to construct the alpha-Voronoi diagrams on the space of discrete probability distributions. Geometry for  $q$ -exponential families, which is related with alpha-geometry, and its statistical applications are also studied. In addition we have studied conformal flatness of level surfaces in Hessian domains. Especially we have also studied harmonic maps between level surfaces of Hessian domains, relating with conformally flat structure.

**Introduction:** Along the line of geometric study of generalized entropies and Legendre structures, we have elucidated a relation between the alpha-geometry and the escort probability, which is an important tool in the arguments of Tsallis's generalized entropy, in the following paper: A. Ohara, H. Matsuzoe and S. Amari, A dually flat structure on the space of escort distributions, 2010 J. Phys.: Conf. Ser. 201 012012 (<http://iopscience.iop.org/1742-6596/201/1/012012>). There we have observed that conformally flattening of the alpha-geometry introduces the escort probabilities as affine coordinates in the resultant dually flat geometry on the space of probability distributions. While this result is still purely mathematical and the implications from viewpoints of statistical physics are necessary, we have found an interesting application to information science.

A  $q$ -exponential family is a set of probability distributions, which is a natural generalization of the standard exponential family, and is related to many physical phenomena called “complex systems” that obey power-laws. A  $q$ -exponential family has geometric structure of constant curvature and a dually flat structure simultaneously. To describe these relations, we introduce a conformal transformation on statistical manifolds and have successfully clarified them in addition to obtaining several important properties. As applications of geometry for  $q$ -exponential families, a geometric generalization of statistical inference are also proposed and studied.

We have also studied Hessian domains, which are flat statistical manifolds typically. It is known that level surfaces of a Hessian domain are 1-conformally flat statistical submanifolds. We showed conditions that 1-conformally flat statistical leaves of a foliation can be realized as level surfaces of their common Hessian domain conversely. In addition we study harmonic maps between level surfaces of a Hessian domain with 1-, (-1)-, and, in general, alpha-conformally flat connections, respectively. Harmonic maps are generalization of critical points of a function, and have been researched in terms of geometry, physics, and so on. For example H. Shima gave conditions for harmonicity of gradient mappings of level surfaces on a Hessian domain. However they investigated harmonic maps on level surfaces into a dual affine space, not into other level surfaces. K. Nomizu and T. Sasaki calculated the Laplacian of centro-affine immersions into an affine space, but we can see no discription of harmonic maps between two centro-affine hypersurfaces. Then we started investigation of harmonic maps between two level surfaces.

**Experiment:** Nothing

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>25 APR 2012</b>		2. REPORT TYPE <b>Final</b>		3. DATES COVERED <b>22-04-2010 to 21-04-2012</b>	
4. TITLE AND SUBTITLE <b>Generalized Entropies and Legendre Duality</b>			5a. CONTRACT NUMBER <b>FA2386-10-1-4060</b>		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) <b>Keiko Uohashi</b>			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Department of Mechanical Engineering &amp; Intelligent Systems, Faculty of Engineeri,Tohoku Gakuin University,1-13-1 Chuo Yagajo,Miyagi Japan,JP,985-8537</b>			8. PERFORMING ORGANIZATION REPORT NUMBER <b>N/A</b>		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) <b>AOARD, UNIT 45002, APO, AP, 96338-5002</b>			10. SPONSOR/MONITOR'S ACRONYM(S) <b>AOARD</b>		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) <b>AOARD-104060</b>		
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT <b>They have shown that a simple and computationally efficient algorithm can be derived to construct the alpha-Voronoi diagrams on the space of discrete probability distributions to make use of conformally flattened structure of alpha-geometry. They also studied 1) geometry for q-exponential families which are related with alpha-geometry, and its statistical applications, and 2) conformal flatness of level surfaces in Hessian domains. Especially they studied harmonic maps between level surfaces of Hessian domains and its relation to conformally flat structure.</b>					
15. SUBJECT TERMS <b>information geometry, generalized entropy, Legendre duality, generalized divergence, Riemannian manifold, affine connection</b>					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>83</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

**Results and Discussion:** We demonstrate that escort probabilities with the new dually flat structure admits a simple algorithm to compute Voronoi diagrams and centroids with respect to alpha-divergences, which are one-parameter distance-like functions representing discrepancy between two probability distributions. The Voronoi diagrams on the space of probability distributions with the Kullback-Leibler, or Bregman divergences have been recognized as important tools for various statistical modeling problems involving pattern classification, clustering, likelihood ratio test and so on [2].

The largest advantage to take account of alpha-divergences is their *invariance under transformations by sufficient statistics* studied by Cencov, which is a significant requirement for those statistical applications. In computational aspect, the conformal flattening of the alpha-geometry enables us to invoke the standard algorithm by Edelsbrunner using a potential function and an upper envelop of hyperplanes with the escort probabilities as coordinates [6].

We elaborate the relations of two structures on  $q$ -exponential family: geometric structure of constant curvature is naturally translated to dually flat structure by conformal transformation. This relation provides us several important geometric properties. One of such examples is a fact that the  $q$ -Pythagorean theorem holds among probability distributions in this family [1]. As a simple application of the theorem, we show that the  $q$ -version of the maximum entropy theorem is naturally induced.

We have also applied obtained mathematical results to extension of statistical inference technique. First we show that the maximizer of the  $q$ -escort distribution is a Bayesian MAP (Maximum A posteriori Probability) estimator [1]. Second, we propose maximum  $q$ -likelihood estimation and geometrically characterize the solution [3].

On conformal flatness of level surfaces in Hessian domains, we obtain the following result [4]. In previous paper we show that a 1-conformally flat statistical manifold can be locally realized as a submanifold of a flat statistical manifold, constructing a level surface of a Hessian domain (Uohashi, Ohara, Fujii; 2000). However we proved realization of only "a" 1-conformally flat statistical manifold. In this study we give conditions for realization of 1-conformally flat statistical manifolds as level surfaces of their common Hessian domain. If embedding a 1-conformally flat statistical model into a higher dimensional model, we may be able to use our result.

To construct harmonic maps, we made mappings from a level surface to another level surface on a Hessian domain by conformal transformation [5]. Next we defined alpha-structure on level surfaces and calculated "variations of mappings" for each alpha-parameters. A harmonic map makes the variation of the mapping zero. So we show a condition for the zero variation by an equation with  $n$  and a parameter "alpha", where  $n$  is dimension of level surfaces. It is a problem to find relations with these harmonic maps and phenomena on statistics, physics, and so on.

**List of Publications:** List any publications, conference presentations, or patents that resulted from this work.

1. Shun-ichi Amari and Atsumi Ohara, Geometry of  $q$ -Exponential Family of Probability Distributions, Entropy, Vol.13, 1170-1185 (2011).
2. Atsumi Ohara, Hiroshi Matsuzoe and Shun-ichi Amari, Conformal geometry of escort probabilities and its application to Voronoi partitions, International Conference on Statistical Physics, Larnaca (Cyprus), July 11-15 2011. Extended version of this paper appeared as Atsumi Ohara, Hiroshi Matsuzoe and Shun-ichi Amari, Dually flat structure with escort probability and its application to alpha-volonoi diagrams, arXiv 1010.4965 (2010)
3. Hiroshi Matsuzoe and Atsumi Ohara, Geometry for  $q$ -exponential families, Recent Progress in Differential Geometry and Its Related Fields: Proceedings of the 2nd International Colloquium on Differential Geometry and its Related Fields, World Scientific Publ., (2011), 55-71.
4. Keiko Uohashi, A Hessian domain constructed with a foliation by 1-conformally flat statistical manifolds, submitted.
5. Keiko Uohashi, Harmonic maps relative to alpha-connections on statistical manifolds, Applied Sciences, Vol.14, 82-88 (2012).

6. Atsumi Ohara, Hiroshi Matsuzoe and Shun-ichi Amari, Conformal Geometry of Escort Probability and its Applications, Modern Physics Letters B, Vol. 26, No. 10, 1250063(14 pages) (2012).

**Attachments:** Publications listed above.

Article

## Geometry of $q$ -Exponential Family of Probability Distributions

Shun-ichi Amari <sup>1,\*</sup> and Atsumi Ohara <sup>2,\*</sup>

<sup>1</sup> Laboratory for Mathematical Neuroscience, RIKEN Brain Science Institute, Hirosawa 2-1, Wako-shi, Saitama 351-0198, Japan

<sup>2</sup> Department of Electrical and Electronics Engineering, Graduate School of Engineering, University of Fukui, Bunkyo 3-9-1, Fukui-shi, Fukui 910-8507, Japan

\* Authors to whom correspondence should be addressed; E-Mails: amari@brain.riken.jp (S.-i.A.); ohara@fuee.u-fukui.ac.jp (A.O.).

Received: 11 February 2011; in revised form: 1 June 2011 / Accepted: 2 June 2011 /

Published: 14 June 2011

---

**Abstract:** The Gibbs distribution of statistical physics is an exponential family of probability distributions, which has a mathematical basis of duality in the form of the Legendre transformation. Recent studies of complex systems have found lots of distributions obeying the power law rather than the standard Gibbs type distributions. The Tsallis  $q$ -entropy is a typical example capturing such phenomena. We treat the  $q$ -Gibbs distribution or the  $q$ -exponential family by generalizing the exponential function to the  $q$ -family of power functions, which is useful for studying various complex or non-standard physical phenomena. We give a new mathematical structure to the  $q$ -exponential family different from those previously given. It has a dually flat geometrical structure derived from the Legendre transformation and the conformal geometry is useful for understanding it. The  $q$ -version of the maximum entropy theorem is naturally induced from the  $q$ -Pythagorean theorem. We also show that the maximizer of the  $q$ -escort distribution is a Bayesian MAP (Maximum A posteriori Probability) estimator.

**Keywords:**  $q$ -exponential family;  $q$ -entropy; information geometry;  $q$ -Pythagorean theorem;  $q$ -Max-Ent theorem; conformal transformation

---

## 1. Introduction

Statistical physics is founded on the Gibbs distribution for microstates, which forms an exponential family of probability distributions known in statistics. Important macro-quantities such as energy, entropy, free energy, *etc.* are connected with it. However, recent studies show that there are non-standard complex systems which are subject to the power law instead of the exponential law of the Gibbs type distributions. See [1,2] as well as extensive literatures cited in them.

Tsallis [3] defined the  $q$ -entropy to elucidate various physical phenomena of this type, followed by many related research works on this subject (see, [1]). The concept of the  $q$ -Gibbs distribution or  $q$ -exponential family of probability distributions is naturally induced from this framework (see also [4]). However, its mathematical structure has not yet been explored enough [2,5,6], while the Gibbs type distribution has been studied well as the exponential family of distributions [7]. We need a mathematical (geometrical) foundation to study the properties of the  $q$ -exponential family. This paper presents a geometrical foundation for the  $q$ -exponential family based on information geometry [8], giving geometrical definitions of the  $q$ -potential function,  $q$ -entropy and  $q$ -divergence in a unified way.

We define the  $q$ -geometrical structure consisting of a Riemannian metric and a pair of dual affine connections. By using this framework, we prove that a family of  $q$ -exponential distributions is dually flat, in which the  $q$ -Pythagorean theorem holds. This naturally induces the corresponding  $q$ -maximum entropy theorem similarly to the case of the Tsallis  $q$ -entropy [1,9,10]. The  $q$ -structure is ubiquitous since the family  $S_n$  of all discrete probability distributions can always be endowed with the structure of the  $q$ -exponential family for arbitrary  $q$ . It is possible to generalize the  $q$ -structure to any family of probability distributions. Further, it has a close relation with the  $\alpha$ -geometry [8], which is one of information geometric structure of constant curvature. This new dually flat structure, different from the old one given rise to from the invariance in information geometry, can be also obtained by conformal flattening of the  $\alpha$ -geometry [11,12], using a technique in the conformal and projective geometry [13–15].

The present framework prepares mathematical tools for analyzing physical phenomena subject to the power law. The Legendre transformation again plays a fundamental role for deriving the geometrical dual structure. There exist lots of applications of  $q$ -geometry to information theory ([16] and others) and statistics, including Bayes  $q$ -statistics.

It is possible to generalize our framework to a more general non-linear family of distributions by using a positive convex function instead of  $q$ -exponential function (See [2,17]). A good example is the  $\kappa$ -exponential family [18–20], but we do not state it here.

## 2. $q$ -Gibbs or $q$ -Exponential Family of Distributions

### 2.1. $q$ -Logarithm and $q$ -Exponential Function

It is the first step to generalize the logarithm and exponential functions to include a family of power functions, where the logarithm and exponential functions are included as the limiting case [1,5,21]. This

was also used for defining the  $\alpha$ -family of distributions in information geometry [8]. We define the  $q$ -logarithm by

$$\log_q(u) = \frac{1}{1-q} (u^{1-q} - 1), \quad u > 0 \quad (1)$$

and its inverse function, the  $q$ -exponential, by

$$\exp_q(u) = \{1 + (1-q)u\}^{\frac{1}{1-q}}, \quad u > -1/(1-q) \quad (2)$$

for a positive  $q$  with  $q \neq 1$ . The limiting case  $q \rightarrow 1$  reduces to

$$\log_1(u) = \log u \quad (3)$$

$$\exp_1(u) = \exp u \quad (4)$$

so that  $\log_q$  and  $\exp_q$  are defined for  $q > 0$ .

## 2.2. $q$ -Exponential Family

The standard form of an exponential family of distributions is written as

$$p(\mathbf{x}, \boldsymbol{\theta}) = \exp \left\{ \sum \theta^i x_i - \psi(\boldsymbol{\theta}) \right\} \quad (5)$$

with respect to an adequate measure  $\mu(\mathbf{x})$ , where  $\mathbf{x} = (x_1, \dots, x_n)$  is a set of random variables and  $\boldsymbol{\theta} = (\theta^1, \dots, \theta^n)$  are the canonical parameters to describe the underlying system. The Gibbs distribution is of this type. Here,  $\psi(\boldsymbol{\theta})$  is called the free energy, which is the cumulant generating function.

The power version of the Gibbs distribution is written as

$$p(\mathbf{x}, \boldsymbol{\theta}) = \exp_q \{ \boldsymbol{\theta} \cdot \mathbf{x} - \psi_q(\boldsymbol{\theta}) \} \quad (6)$$

$$\log_q \{ p(\mathbf{x}, \boldsymbol{\theta}) \} = \boldsymbol{\theta} \cdot \mathbf{x} - \psi_q(\boldsymbol{\theta}) \quad (7)$$

where  $\boldsymbol{\theta} \cdot \mathbf{x} = \sum \theta^i x_i$ . This is the  $q$ -Gibbs distribution or  $q$ -exponential family [4], which we denote by  $S$ , where the domain of  $\mathbf{x}$  is restricted such that  $p(\mathbf{x}, \boldsymbol{\theta}) > 0$  holds. The function  $\psi_q(\boldsymbol{\theta})$ , called the  $q$ -free energy or  $q$ -potential function, is determined from the normalization condition:

$$\int \exp_q \{ \boldsymbol{\theta} \cdot \mathbf{x} - \psi_q(\boldsymbol{\theta}) \} d\mathbf{x} = 1 \quad (8)$$

where we replaced  $d\mu(\mathbf{x})$  by  $d\mathbf{x}$  for brevity's sake. The function  $\psi_q$  depends on  $q$ , but we hereafter neglect suffix  $q$  in most cases. Research on the  $q$ -exponential family can be found, for example, in [2,4,19]. The  $q$ -Gaussian distribution is given by

$$p(x, \mu, \sigma) = \exp_q \left\{ -\frac{(x - \mu)^2}{2\sigma^2} - \psi(\mu, \sigma) \right\} \quad (9)$$

and is studied in [22–25] in detail. Here, we need to introduce a vector random variable  $\mathbf{x} = (x, x^2)$  and a new parameter  $\boldsymbol{\theta}$ , which is a vector-valued function of  $\mu$  and  $\sigma$ , to represent it in the standard form (7). It is an interesting observation that the domain of  $\mathbf{x}$  in the  $q$ -Gaussian case depends on  $q$  if  $0 < q < 1$ . Hence, that  $q$ - and  $q'$ -Gaussian are in general not absolutely continuous when  $q \neq q'$ .

It should be remarked that the  $q$ -exponential family itself is the same as the  $\alpha$ -family of distributions in information geometry [8]. Here, we introduce a different geometrical structure, generalizing the result of [24].

We mainly use the family  $S_n$  of discrete distributions over  $(n + 1)$  elements  $X = \{x_0, x_1, \dots, x_n\}$ , although we can easily extend the results to the case of continuous random variables. Here, random variable  $x$  takes values over  $X$ . We also treat the case of  $0 < q < 1$ , and the limiting cases of  $q = 0$  or  $1$  give the well-known ones.

Let us put  $p_i = \text{Prob} \{x = x_i\}$  and denote the probability distribution by vector  $\mathbf{p} = (p_0, p_1, \dots, p_n)$ , where

$$\sum_{i=1}^n p_i = 1 \quad (10)$$

The probability of  $x$  is also written as

$$p(x) = \sum_{i=0}^n p_i \delta_i(x) \quad (11)$$

where

$$\delta_i(x) = \begin{cases} 1, & x = x_i, \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

**Theorem 1** The family  $S_n$  of discrete probability distributions has the structure of a  $q$ -exponential family for any  $q$ .

*Proof* We take  $\log_q$  of distribution  $p(x)$  of (11). For any function  $f(u)$ , we have

$$f \left\{ \sum_{i=1}^n p_i \delta_i(x) \right\} = \sum_{i=0}^n f(p_i) \delta_i(x) \quad (13)$$

By taking

$$\delta_0(x) = 1 - \sum_{i=1}^n \delta_i(x) \quad (14)$$

into account, discrete distribution (11) can be rewritten in the form (8) as

$$\log_q p(x) = \frac{1}{1-q} \left\{ \sum_{i=1}^n (p_i^{1-q} - p_0^{1-q}) \delta_i(x) + p_0^{1-q} - 1 \right\} \quad (15)$$

where

$$p_0 = 1 - \sum_{i=1}^n p_i \quad (16)$$

is treated as a function of  $(p_1, \dots, p_n)$ . Hence,  $S_n$  is  $q$ -exponential family (6) for any  $q$ , with the following  $q$ -canonical parameters, random variables and  $q$ -potential function:

$$\theta^i = \frac{1}{1-q} (p_i^{1-q} - p_0^{1-q}), \quad i = 1, \dots, n \quad (17)$$

$$x_i = \delta_i(x) \quad (18)$$

$$\psi(\boldsymbol{\theta}) = -\log_q p_0 \quad (19)$$



This completes the proof.  $\square$

Note that the  $q$ -potential  $\psi(\boldsymbol{\theta})$  and the canonical parameter  $\boldsymbol{\theta}$  depend on  $q$  as is seen in (17) and (19). It should also be remarked that Theorem 1 does not contradict to the theorem 1 in [19] stating that a parametrized family of probability distributions can belong to at most one  $q$ -exponential family. The author considers an  $m$ -dimensional parametrized submanifold in  $S_n$  with  $m < n$  where the canonical parameter depending on  $q$  is given via the variational principle. Therefore, by denoting the  $q$ -canonical parameter by  $\boldsymbol{\theta}_q \in \mathbb{R}^m$ , we can restate his theorem in terms of geometry that a linear submanifold parametrized by  $\boldsymbol{\theta}_q \in \mathbb{R}^m$  is not a linear submanifold parametrized by  $\boldsymbol{\theta}_{q'} \in \mathbb{R}^m$  when  $q' \neq q$ . On the other hand, the present theorem states that there exists the  $q$ -canonical parameter  $\boldsymbol{\theta}_q \in \mathbb{R}^n$  on whole  $S_n$  for any  $q$  and the manifold has linear structure with respect to any  $\boldsymbol{\theta}_q$ . This is a surprising new finding.

### 2.3. $q$ -Potential Function

We study the  $q$ -geometrical structure of  $S$ . The  $q$ -log-likelihood is a linear form defined by

$$l_q(\mathbf{x}, \boldsymbol{\theta}) = \log_q p(\mathbf{x}, \boldsymbol{\theta}) = \sum_{i=1}^n \theta^i x_i - \psi(\boldsymbol{\theta}) \quad (20)$$

By differentiating it with respect to  $\theta^i$ , with the abbreviated notation  $\partial_i = \frac{\partial}{\partial \theta^i}$ , we have

$$\partial_i l_q(\mathbf{x}, \boldsymbol{\theta}) = x_i - \partial_i \psi(\boldsymbol{\theta}) \quad (21)$$

$$\partial_i \partial_j l_q(\mathbf{x}, \boldsymbol{\theta}) = -\partial_i \partial_j \psi(\boldsymbol{\theta}) \quad (22)$$

From this we have the following important theorem.

**Theorem 2** The  $q$ -free energy or  $q$ -potential  $\psi_q(\boldsymbol{\theta})$  is a convex function of  $\boldsymbol{\theta}_q$ .

*Proof* We omit the suffix  $q$  for simplicity's sake. We have

$$\partial_i p(\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{x}, \boldsymbol{\theta})^q (x_i - \partial_i \psi) \quad (23)$$

$$\partial_i \partial_j p(\mathbf{x}, \boldsymbol{\theta}) = qp(\mathbf{x}, \boldsymbol{\theta})^{2q-1} (x_i - \partial_i \psi)(x_j - \partial_j \psi) - p(\mathbf{x}, \boldsymbol{\theta})^q \partial_i \partial_j \psi \quad (24)$$

The following identities hold:

$$\int \partial_i p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} = \partial_i \int p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} = 0 \quad (25)$$

$$\int \partial_i \partial_j p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} = \partial_i \partial_j \int p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} = 0 \quad (26)$$

Here, we define an important functional

$$h_q(\boldsymbol{\theta}) = h_q[p(\mathbf{x}, \boldsymbol{\theta})] = \int p(\mathbf{x}, \boldsymbol{\theta})^q d\mathbf{x} \quad (27)$$

in particular for discrete  $S_n$ ,

$$h_q(\mathbf{p}) = \sum_{i=0}^n p_i^q \quad (28)$$

for  $0 < q < 1$ . This function plays a key role in the following. From (25) and (26), by using (23) and (24), we have

$$\partial_i \psi(\boldsymbol{\theta}) = \frac{1}{h_q(\boldsymbol{\theta})} \int x_i p(\mathbf{x}, \boldsymbol{\theta})^q d\mathbf{x} \quad (29)$$

$$\partial_i \partial_j \psi(\boldsymbol{\theta}) = \frac{q}{h_q(\boldsymbol{\theta})} \int (x_i - \partial_i \psi)(x_j - \partial_j \psi) p(\mathbf{x}, \boldsymbol{\theta})^{2q-1} d\mathbf{x} \quad (30)$$

The latter shows that  $\partial_i \partial_j \psi(\boldsymbol{\theta})$  is positive-definite, and hence  $\psi$  is convex.  $\square$

#### 2.4. $q$ -Divergence

A convex function  $\psi(\boldsymbol{\theta})$  makes it possible to define a divergence of the Bregman-type between two probability distributions  $p(\mathbf{x}, \boldsymbol{\theta}_1)$  and  $p(\mathbf{x}, \boldsymbol{\theta}_2)$  [8,26,27]. It is given by using the gradient  $\nabla = \partial/\partial\boldsymbol{\theta}$ ,

$$D_q[p(\mathbf{x}, \boldsymbol{\theta}_1) : p(\mathbf{x}, \boldsymbol{\theta}_2)] = \psi(\boldsymbol{\theta}_2) - \psi(\boldsymbol{\theta}_1) - \nabla \psi(\boldsymbol{\theta}_1) \cdot (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1) \quad (31)$$

satisfying the non-negativity condition

$$D_q[p(\mathbf{x}, \boldsymbol{\theta}_1) : p(\mathbf{x}, \boldsymbol{\theta}_2)] \geq 0 \quad (32)$$

with equality when and only when  $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ . This gives a  $q$ -divergence in  $S_n$  different from the invariant divergence of  $S_n$  [28]. The divergence is canonical in the sense that it is uniquely determined in accordance with dually flat structure of  $q$ -exponential family in Sections 3 and 4. The canonical divergence is different from the  $\alpha$ -divergence or conventional Tsallis relative entropy used in information geometry (See the discussion in the end of this subsection). Note that it is used in [16].

**Theorem 3** For two discrete distributions  $p(x) = \mathbf{p}$  and  $r(x) = \mathbf{r}$ , the  $q$ -divergence is given by

$$D_q[\mathbf{p} : \mathbf{r}] = \frac{1}{(1-q)h_q(\mathbf{p})} \left( 1 - \sum_{i=0}^n p_i^q r_i^{1-q} \right) \quad (33)$$

*Proof* The potentials are, from (19),

$$\psi(\mathbf{p}) = -\log_q p_0, \quad \psi(\mathbf{r}) = -\log_q r_0 \quad (34)$$

for  $\mathbf{p}$  and  $\mathbf{r}$ . We need to calculate  $\nabla \psi(\boldsymbol{\theta})$  given in (29). In our case,  $x_i = \delta_i(x)$  and hence

$$\partial_i \psi = \frac{p_i^q}{h_q(\mathbf{p})} \quad (35)$$

By using this and (17), we obtain (33).  $\square$

It is useful to consider a related probability distribution,

$$\hat{p}_q(\mathbf{x}) = \frac{1}{h_q[p(\mathbf{x})]} p(\mathbf{x})^q \quad (36)$$

for defining the  $q$ -expectation. This is called the  $q$ -escort probability distribution [1,4,29]. Introducing the  $q$ -expectation of random variable  $f(\mathbf{x})$  by

$$E_{\hat{p}}[f(\mathbf{x})] = \frac{1}{h_q[p(\mathbf{x})]} \int p(\mathbf{x})^q f(\mathbf{x}) d\mathbf{x} \quad (37)$$

we can rewrite the  $q$ -divergence (31) for  $p(\mathbf{x}), r(\mathbf{x}) \in S$  as

$$D_q[p(\mathbf{x}) : r(\mathbf{x})] = E_{\hat{p}}[\log_q p(\mathbf{x}) - \log_q r(\mathbf{x})] \quad (38)$$

because of the relations (20) and (29). The expression (38) is also valid on the exterior of  $S \times S$  when it is integrable. This is different from the definition of the Tsallis relative entropy [30,31]

$$\tilde{D}_q[p(\mathbf{x}) : r(\mathbf{x})] = \frac{1}{1-q} \left( 1 - \int p(\mathbf{x})^q r(\mathbf{x})^{1-q} d\mathbf{x} \right) \quad (39)$$

which is equal to the well-known  $\alpha$ -divergence up to a constant factor where  $\alpha = 1 - 2q$  (see [8,28]), satisfying the invariance criterion. We have

$$D_q[p(\mathbf{x}) : r(\mathbf{x})] = \frac{1}{h_q[p(\mathbf{x})]} \tilde{D}_q[p(\mathbf{x}) : r(\mathbf{x})] \quad (40)$$

This is a conformal transformation of divergence, as we see in the following. See also the derivation based on affine differential geometry [12].

## 2.5. $q$ -Riemannian Metric

When  $\theta_2$  is infinitesimally close to  $\theta_1$ , by putting  $\theta_1 = \theta$ ,  $\theta_2 = \theta + d\theta$  and using the Taylor expansion, we have

$$D_q[p(\mathbf{x}, \theta) : p(\mathbf{x}, \theta + d\theta)] = \sum g_{ij}^q(\theta) d\theta^i d\theta^j \quad (41)$$

where

$$g_{ij}^{(q)} = \partial_i \partial_j \psi(\theta) \quad (42)$$

is a positive-definite matrix. We call  $[g_{ij}^{(q)}(\theta)]$  the  $q$ -Fisher information matrix. When  $q = 1$ , this reduces to the ordinary Fisher information matrix given by

$$g_{ij}^{(1)}(\theta) = g_{ij}^F(\theta) = E[\partial_i \log p(\mathbf{x}, \theta) \partial_j \log p(\mathbf{x}, \theta)] \quad (43)$$

The positive-definite matrix  $g_{ij}^{(q)}(\theta)$  defines a Riemannian metric on  $S_n$ , giving it the  $q$ -Riemannian structure.

When a metric tensor  $g_{ij}(\theta)$  is transformed to

$$\tilde{g}_{ij}(\theta) = \sigma(\theta) g_{ij}(\theta) \quad (44)$$

by a positive function  $\sigma(\theta)$ , we call it a conformal transformation. See, e.g., [13–15,32]. The conformal transformation of divergence induces that of the Riemannian metric.

**Theorem 4** The  $q$ -Fisher information metric is given by a conformal transformation of the Fisher information metric  $g_{ij}^F$  as

$$g_{ij}^{(q)}(\boldsymbol{\theta}) = \frac{q}{h_q(\boldsymbol{\theta})} g_{ij}^F(\boldsymbol{\theta}) \quad (45)$$

*Proof* The  $q$ -metric is derived from the Taylor expansion of  $D_q[p : p + d\mathbf{p}]$ . We have

$$\begin{aligned} D_q[p(\mathbf{x}, \boldsymbol{\theta}) : p(\mathbf{x}, \boldsymbol{\theta} + d\boldsymbol{\theta})] &= \frac{1}{(1-q)h_q(\boldsymbol{\theta})} \left\{ 1 - \int p(\mathbf{x}, \boldsymbol{\theta})^q p(\mathbf{x}, \boldsymbol{\theta} + d\boldsymbol{\theta})^{1-q} d\mathbf{x} \right\} \\ &= \frac{q}{h_q(\boldsymbol{\theta})} \left\{ \int \frac{1}{p(\mathbf{x}, \boldsymbol{\theta})} \partial_i p(\mathbf{x}, \boldsymbol{\theta}) \partial_j p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} \right\} d\theta^i d\theta^j \end{aligned} \quad (46)$$

using the identities (25) and (26). When  $q = 1$ , this is the Fisher information given by (43). Hence, the  $q$ -Fisher information is given by (45).  $\square$

A Riemannian metric defines the length of a tangent vector  $\mathbf{X} = (X^1, \dots, X^n)$  at  $\boldsymbol{\theta}$  by

$$\|\mathbf{X}\|^2 = \sum g_{ij}(\boldsymbol{\theta}) X^i X^j \quad (47)$$

Similarly, for two tangent vectors  $\mathbf{X}$  and  $\mathbf{Y}$ , their inner product is defined by

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum g_{ij} X^i Y^j \quad (48)$$

When  $\langle \mathbf{X}, \mathbf{Y} \rangle$  vanishes,  $\mathbf{X}$  and  $\mathbf{Y}$  are said to be orthogonal. The orthogonality, or more generally the angle, of two vectors  $\mathbf{X}$  and  $\mathbf{Y}$  does not change by a conformal transformation, although their magnitudes change.

### 3. Dually Flat Structure of $q$ -Exponential Family

#### 3.1. Legendre Transformation and $q$ -Entropy

Given a convex function  $\psi(\boldsymbol{\theta})$ , the Legendre transformation is defined by

$$\boldsymbol{\eta} = \nabla \psi(\boldsymbol{\theta}) \quad (49)$$

where  $\nabla = (\partial/\partial\theta^i)$  is the gradient. Since the correspondence between  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  is one-to-one, we may consider  $\boldsymbol{\eta}$  as another coordinate system of  $S$ .

The dual potential function is defined by

$$\varphi(\boldsymbol{\eta}) = \max_{\boldsymbol{\theta}} \{ \boldsymbol{\theta} \cdot \boldsymbol{\eta} - \psi(\boldsymbol{\theta}) \} \quad (50)$$

which is convex with respect to  $\boldsymbol{\eta}$ . The original coordinates are recovered from the inverse transformation given by

$$\boldsymbol{\theta} = \nabla \varphi(\boldsymbol{\eta}) \quad (51)$$

where  $\nabla = (\partial/\partial\eta_i)$ , so that  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  are in dual correspondence.

The following theorem gives explicit relations among these quantities.

**Theorem 5** The dual coordinates  $\boldsymbol{\eta}$  are given by

$$\boldsymbol{\eta} = E_{\hat{p}}[\mathbf{x}] \quad (52)$$

and the dual potential is given by

$$\varphi(\boldsymbol{\eta}) = \frac{1}{1-q} \left\{ \frac{1}{h_q(\mathbf{p})} - 1 \right\} \quad (53)$$

*Proof* The relation (52) is immediate from (29). From the Legendre duality, the dual potential satisfies

$$\varphi(\boldsymbol{\eta}) + \psi(\boldsymbol{\theta}) - \boldsymbol{\theta} \cdot \boldsymbol{\eta} = 0 \quad (54)$$

when  $\boldsymbol{\theta}$  and  $\boldsymbol{\eta}$  correspond to each other by  $\boldsymbol{\eta} = \nabla \psi(\boldsymbol{\theta})$ . Therefore,

$$\varphi(\boldsymbol{\eta}) = \sum_{i=1}^n \theta^i \eta_i - \psi(\boldsymbol{\theta}) \quad (55)$$

$$= E_{\hat{p}} [\log_q p(\mathbf{x}, \boldsymbol{\theta})] \quad (56)$$

$$= \frac{1}{(1-q)h_q(\boldsymbol{\theta})} \left( 1 - \int p^q(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} \right) \quad (57)$$

$$= \frac{1}{1-q} \left( \frac{1}{h_q(\boldsymbol{\theta})} - 1 \right) \quad (58)$$

This is a convex function of  $\boldsymbol{\eta}$ .  $\square$

We call the  $q$ -dual potential

$$\varphi(\boldsymbol{\eta}) = E [\log_q p(\mathbf{x}, \boldsymbol{\theta})] = \frac{1}{1-q} \left\{ \frac{1}{h_q} - 1 \right\} \quad (59)$$

the negative  $q$ -entropy, because it is the Legendre-dual of the  $q$ -free energy  $\psi(\boldsymbol{\theta})$ . There are various definitions of  $q$ -entropy. The Tsallis  $q$ -entropy [3] is originally defined by

$$H_{\text{Tsallis}} = \frac{1}{1-q} (h_q - 1) \quad (60)$$

while the Rényi  $q$ -entropy [33] is

$$H_{\text{Rényi}} = \frac{1}{1-q} \log h_q \quad (61)$$

They are mutually related by monotone functions. When  $q \rightarrow 1$ , all of them reduce to the Shannon entropy.

Our definition of

$$H_q = \frac{1}{1-q} \left( 1 - \frac{1}{h_q} \right) = \frac{H_{\text{Tsallis}}}{h_q} \quad (62)$$

is also monotonically connected with the previous ones, but is more natural from the point of view of  $q$ -geometry. The entropy  $H_q$  has been known as the normalized  $q$ -entropy, which was studied in [16,34–37].

### 3.2. $q$ -Dually Flat Structure

There are two dually coupled coordinate systems  $\theta$  and  $\eta$  in  $q$ -exponential family  $S$  with two potential functions  $\psi(\theta)$  and  $\varphi(\eta)$  for each  $q$ . Two affine structures are introduced by the two convex functions  $\psi$  and  $\varphi$ . See information geometry of dually flat space [8]. Although  $S$  is a Riemannian manifold given by the  $q$ -Fisher information matrix (45), we may nevertheless regard  $S$  as an affine manifold where  $\theta$  is an affine coordinate system. They represent intensive quantities of a physical system. Dually, we introduce a dual affine structure to  $S$ , where  $\eta$  is another affine coordinate system. They represent extensive quantities. We can define two types of straight lines or geodesics in  $S$  due to the  $q$ -affine structures.

For two distributions  $p(\mathbf{x}, \theta_1)$  and  $p(\mathbf{x}, \theta_2)$  in  $S$ , a curve  $p(\mathbf{x}, \theta(t))$  is said to be a  $q$ -geodesic connecting them, when

$$\theta(t) = t\theta_1 + (1 - t)\theta_2 \quad (63)$$

where  $t$  is the parameter of the curve. Dually, in terms of dual coordinates  $\eta$ , when

$$\eta(t) = t\eta_1 + (1 - t)\eta_2 \quad (64)$$

holds, the curve is said to be a dual  $q$ -geodesic.

More generally, the  $q$ -geodesic connecting two distribution  $p_1(\mathbf{x})$  and  $p_2(\mathbf{x})$  is given by

$$\log_q p(\mathbf{x}, t) = t \log_q p_1(\mathbf{x}) + (1 - t) \log_q p_2(\mathbf{x}) - c(t) \quad (65)$$

where  $c(t)$  is a normalizing term. This is rewritten as

$$p(\mathbf{x}, t)^{1-q} = tp_1(\mathbf{x})^{1-q} + (1 - t)p_2(\mathbf{x})^{1-q} - c(t) \quad (66)$$

Dually, the dual  $q$ -geodesic connecting  $p_1(\mathbf{x})$  and  $p_2(\mathbf{x})$  is given by using the escort distributions as

$$\hat{p}(\mathbf{x}, t) = t\hat{p}_1(\mathbf{x}) + (1 - t)\hat{p}_2(\mathbf{x}) \quad (67)$$

Since the manifold  $S$  has a  $q$ -Riemannian structure, the orthogonality of two tangent vectors is defined by the Riemannian metric. We rewrite the orthogonality of two geodesics in terms of the affine coordinates. Let us consider two small deviations  $d_1 p(\mathbf{x})$  and  $d_2 p(\mathbf{x})$  of  $p(\mathbf{x})$ , that is, from  $p(\mathbf{x})$  to  $p(\mathbf{x}) + d_1 p(\mathbf{x})$  and  $p(\mathbf{x}) + d_2 p(\mathbf{x})$ , which are regarded as two (infinitesimal) tangent vectors of  $S$  at  $p(\mathbf{x})$ .

**Lemma 1** The inner product of two deviations  $d_1 p$  and  $d_2 p$  is given by

$$\langle d_1 p(\mathbf{x}), d_2 p(\mathbf{x}) \rangle = \int d_1 \hat{p}(\mathbf{x}) d_2 \log_q p(\mathbf{x}) d\mathbf{x} \quad (68)$$

*Proof* By simple calculations, we have

$$\int d_1 \hat{p}(\mathbf{x}) d_2 \log_q p(\mathbf{x}) d\mathbf{x} = \frac{q}{h_q} \int \frac{d_1 p(\mathbf{x}) d_2 p(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x} \quad (69)$$

of which the right-hand side is the Riemannian inner product in the form of (46).  $\square$

**Corollary.** Two curves  $\theta_1(t)$  and  $\eta_2(t)$ , intersecting at  $t = 0$ , are orthogonal when  $\langle \dot{\theta}_1(0), \dot{\eta}_2(0) \rangle = 0$ . Here,  $\dot{\theta}_1(t)$  and  $\dot{\eta}_2(t)$  denote derivatives of  $\theta_1(t)$  and  $\eta_2(t)$  by  $t$ , respectively.

The two geodesics and the orthogonality play a fundamental role in  $S$  as will be seen in the following.

#### 4. $q$ -Pythagorean and $q$ -Max-Ent Theorems

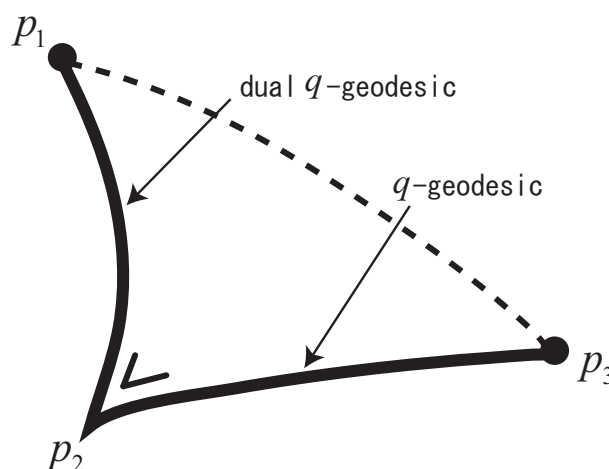
A dually flat Riemannian manifold admits the generalized Pythagorean theorem and the related projection theorem [8]. We state them in our case.

**$q$ -Pythagorean Theorem.** For three distributions  $p_1(\mathbf{x})$ ,  $p_2(\mathbf{x})$  and  $p_3(\mathbf{x})$  in  $S$ , it holds that

$$D_q[p_1 : p_2] + D_q[p_2 : p_3] = D_q[p_1 : p_3] \quad (70)$$

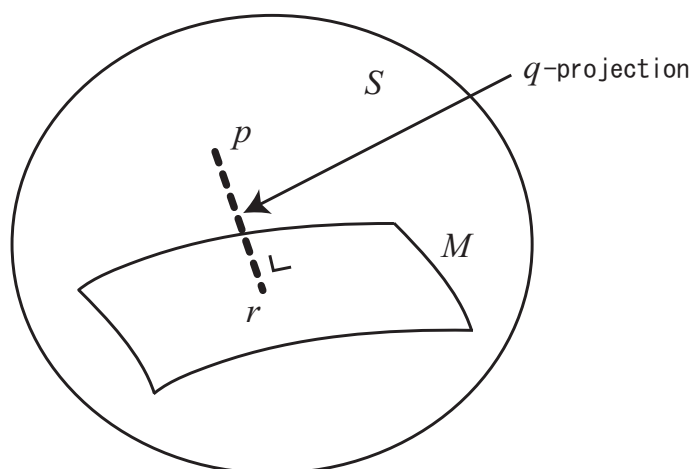
when the dual geodesic connecting  $p_1(\mathbf{x})$  and  $p_2(\mathbf{x})$  is orthogonal at  $p_2(\mathbf{x})$  to the geodesic connecting  $p_2(\mathbf{x})$  and  $p_3(\mathbf{x})$  (see Figure 1).

**Figure 1.**  $q$ -Pythagorean theorem.



Given a distribution  $p(\mathbf{x}) \in S$  and a submanifold  $M \subset S$ , a distribution  $r(\mathbf{x}) \in M$  is said to be the  $q$ -projection (dual  $q$ -projection) of  $p(\mathbf{x})$  to  $M$ , when the  $q$ -geodesic (dual  $q$ -geodesic) connecting  $p(\mathbf{x})$  and  $r(\mathbf{x})$  is orthogonal to  $M$  at  $r(\mathbf{x})$  (Figure 2).

**Figure 2.**  $q$ -projection of  $p$  to  $M$ .



**$q$ -Projection Theorem.** Let  $M$  be a submanifold of  $S$ . Given  $p(\mathbf{x}) \in S$ , the point  $r(\mathbf{x}) \in M$  that minimizes  $D_q[p(\mathbf{x}) : r(\mathbf{x})]$  is given by the dual  $q$ -projection of  $p(\mathbf{x})$  to  $M$ . The point  $r(\mathbf{x}) \in M$  that minimizes  $D_q[r(\mathbf{x}) : p(\mathbf{x})]$  is given by the  $q$ -projection of  $p(\mathbf{x})$  to  $M$ .

We show that the well-known  $q$ -max-ent theorem in the case of Tsallis  $q$ -entropy [1,4,9,11] is a direct consequence of the above  $q$ -Pythagorean and  $q$ -projection theorems.

**$q$ -Max-Ent Theorem.** Probability distributions maximizing the  $q$ -entropies  $H_{\text{Tsallis}}$ ,  $H_{\text{Rényi}}$  and  $H_q$  under  $q$ -linear constraints for  $m$  random variables  $c_k(\mathbf{x})$  and various values of  $a_k$

$$E_{\hat{p}}[c_k(\mathbf{x})] = a_k, \quad k = 1, \dots, m \quad (71)$$

form a  $q$ -exponential family

$$\log_q p(\mathbf{x}, \boldsymbol{\theta}) = \sum_{i=1}^m \theta^i c_i(\mathbf{x}) - \psi(\boldsymbol{\theta}) \quad (72)$$

The proof is easily obtained by the standard analytical method. Here, we give a geometrical proof. Let us consider the subspace  $M^* \subset S$  whose member  $p(\mathbf{x})$  satisfies the  $m$  constraints

$$E_{\hat{p}}[c_k(\mathbf{x})] = \int \hat{p}(\mathbf{x}) c_k(\mathbf{x}) d\mathbf{x} = a_k, \quad k = 1, \dots, m. \quad (73)$$

Since the constraints are linear in the dual affine coordinates  $\boldsymbol{\eta}$  or  $\hat{p}(\mathbf{x})$ ,  $M^*$  is a linear subspace of  $S$  with respect to the dual affine connection. Let  $p_0(\mathbf{x}, \boldsymbol{\theta}_0)$  be the uniform distribution defined by  $\boldsymbol{\theta}_0 = 0$ , which implies  $p_0(\mathbf{x}, \boldsymbol{\theta}_0) = \text{const}$  from (6). Let  $\bar{p}(\mathbf{x}) \in M^*$  be the  $q$ -projection of  $p_0(\mathbf{x})$  to  $M^*$  (Figure 3). Then, the divergence  $D_q[p : p_0]$  from  $p(\mathbf{x}) \in M^*$  to  $p_0(\mathbf{x})$  is decomposed as

$$D_q[p : p_0] = D_q[p : \bar{p}] + D_q[\bar{p} : p_0] \quad (74)$$

Let  $\boldsymbol{\eta}_p$  be the dual coordinates of  $p(\mathbf{x})$ . Since the divergence is written as

$$D_q[p : p_0] = \psi(\boldsymbol{\theta}_0) + \varphi(\boldsymbol{\eta}_p) - \boldsymbol{\theta}_0 \cdot \boldsymbol{\eta}_p \quad (75)$$

the minimizer of  $D_q[p : p_0]$  among  $p(\mathbf{x}) \in M^*$  is just  $\bar{p}(\mathbf{x})$ , which is also the maximizer of the entropy  $-\varphi(\boldsymbol{\eta}_p)$ .

The trajectories of  $\bar{p}(\mathbf{x})$  for various values of  $a_k$  form a flat subspace orthogonal to  $M^*$ , implying that they form a  $q$ -exponential family of the form (6) (see Figure 3). The tangent directions  $d\hat{p}(\mathbf{x})$  of  $M^*$  satisfies

$$\int d\hat{p}(\mathbf{x}) c_k(\mathbf{x}) d\mathbf{x} = 0, \quad k = 1, \dots, m. \quad (76)$$

Hence, a  $q$ -exponential family of the form

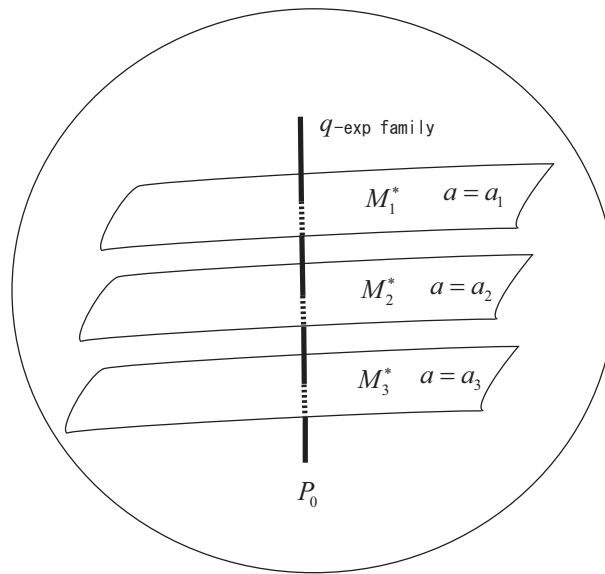
$$\log_q p(\mathbf{x}, \boldsymbol{\xi}) = \sum_{i=1}^m \xi_i d_i(\mathbf{x}) - \psi(\boldsymbol{\xi}) \quad (77)$$

is orthogonal to  $M^*$ , when

$$\int d\hat{p}(\mathbf{x}) d \log_q p(\mathbf{x}, \boldsymbol{\xi}) d\mathbf{x} = 0 \quad (78)$$

This implies that  $d_i(\mathbf{x}) = c_i(\mathbf{x})$ . Hence, we have the  $q$ -exponential family (72) that maximizes the  $q$ -entropies.



**Figure 3.**  $q$ -Max-Ent theorem.

### 5. $q$ -Bayesian MAP Estimator

Given  $N$  iid observations  $\mathbf{x}_1, \dots, \mathbf{x}_N$  from a statistical model  $M = \{p(\mathbf{x}, \boldsymbol{\xi})\}$ , we have

$$p(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\xi}) = \prod_{i=1}^N p(\mathbf{x}_i, \boldsymbol{\xi}) \quad (79)$$

Since  $\log_q u$  is a monotonically increasing function, the maximizer of the  $q$ -likelihood

$$l_q(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\xi}) = \log_q p(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\xi}) \quad (80)$$

is the same as the ordinary maximum likelihood estimator (mle). However, the maximizer of the  $q$ -escort distribution that maximizes the  $q$ -escort log-likelihood,

$$\frac{1}{q} \hat{l}(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\xi}) = \log p(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\xi}) - \frac{1}{q} \log h_q(\boldsymbol{\xi}) \quad (81)$$

is different from this. We show that the  $q$ -mle is a Bayesian MAP (maximum a posteriori probability) estimator. This clarifies the meaning of the  $q$ -escort mle.

The  $q$ -escort mle is the maximizer of the  $q$ -escort distribution,

$$\hat{\boldsymbol{\xi}}_q = \arg \max \hat{p}(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\xi}) \quad (82)$$

**Theorem 6** The  $q$ -escort mle  $\hat{\boldsymbol{\xi}}_q$  is the Bayesian MAP estimator with the prior distribution

$$\pi(\boldsymbol{\xi}) = h_q(\boldsymbol{\xi})^{-N/q} \quad (83)$$

*Proof* The Bayesian MAP is the maximizer of the posterior distribution with prior  $\pi(\boldsymbol{\xi})$

$$p(\boldsymbol{\xi} | \mathbf{x}_1, \dots, \mathbf{x}_N) = \frac{\pi(\boldsymbol{\xi}) p(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\xi})}{p(\mathbf{x}_1, \dots, \mathbf{x}_N)} \quad (84)$$

which also maximizes

$$(\pi(\boldsymbol{\xi})p(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\xi}))^q, \text{ for } q > 0 \quad (85)$$

On the other hand, the  $q$ -escort mle is the maximizer of

$$\hat{p}(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\xi}) = \prod_{i=1}^N \hat{p}(\mathbf{x}_i, \boldsymbol{\xi}) = \prod_{i=1}^N \frac{p(\mathbf{x}_i, \boldsymbol{\xi})^q}{h_q(\boldsymbol{\xi})} \quad (86)$$

Hence, when

$$\pi(\boldsymbol{\xi}) = h_q(\boldsymbol{\xi})^{-N/q} \quad (87)$$

the two estimators are identical.  $\square$

The theorem shows that the Bayesian prior has a peak at the maximizer of our  $q$ -entropy  $H_q$ .

## 6. Conclusions

Much attention has been recently paid to the probability distributions subject to the power law, instead of the exponential law, since Tsallis proposed the  $q$ -entropy and related theories. The power law is also found in various communication networks. It is now a hot topic of research.

However, we do not have a geometrical foundation while that for the ordinary family of probability distributions is given by information geometry [8]. The present paper tried to give a geometrical foundation to the  $q$ -family of probability distributions. We introduced a new notion of the  $q$ -geometry. The  $q$ -structure is ubiquitous in the sense that the family of all the discrete probability distributions (and the family of all the continuous probability distributions, if we neglect delicate problems involved in the infinite dimensionality) belongs to the  $q$ -exponential family of distributions for any  $q$ . That is, we can introduce the  $q$ -geometrical structure to an arbitrary family of probability distributions, because any parametrized family of probability distributions forms a submanifold embedded in the entire manifold.

The  $q$ -structure consists of a Riemannian metric together with a pair of dually coupled affine connections, which sits in the framework of the standard information geometry. However, the  $q$ -structure is essentially different from the standard one derived by the invariance criterion of the manifold of probability distributions. We have a novel look on the theory related to the  $q$ -entropy from a viewpoint of conformal transformation. This leads us to unified definitions of various quantities such as the  $q$ -entropy,  $q$ -divergence,  $q$ -potential function and their duals, as well as new interpretations of known quantities.

This is a geometrical foundation and we expect that the paper contributes to provide further developments in this field.

## References

1. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics*; Springer: New York, NY, USA, 2009.
2. Naudts, J. *Generalised Thermostatistics*; Springer: London, UK, 2011.
3. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487.
4. Naudts, J. The  $q$ -exponential family in statistical Physics. *Cent. Eur. J. Phys.* **2009**, *7*, 405–413.

5. Suyari, H. Mathematical structures derived from the  $q$ -multinomial coefficient in Tsallis statistics. *Physica A* **2006**, *368*, 63–82.
6. Suyari, H.; Wada, T. Multiplicative duality,  $q$ -triplet and  $\mu, \nu, q$ -relation derived from the one-to-one correspondence between the  $(\mu, \nu)$ -multinomial coefficient and Tsallis entropy  $S_q$ . *Physica A* **2008**, *387*, 71–83.
7. Barndorff-Nielsen, O.E. *Information and Exponential Families in Statistical Theory*. Wiley: New York, NY, USA, 1978.
8. Amari, S.; Nagaoka, H. *Methods of Information Geometry (Translations of Mathematical Monographs)*; Oxford University Press: Oxford, UK, 2000.
9. Ohara, A. Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization. *Phys. Lett. A* **2007**, *370*, 184–193.
10. Furuichi, S. On the maximum entropy principle and the minimization of the Fisher information in Tsallis statistics. *J. Math. Phys.* **2009** *50*, 013303.
11. Ohara, A. Geometric study for the Legendre duality of generalized entropies and its application to the porous medium equation. *Eur. Phys. J. B* **2009**, *70*, 15–28.
12. Ohara, A.; Matsuzoe, H.; Amari, S. A dually flat structure with escort probability and its application to alpha-Voronoi diagrams. *arXiv* **2010**, arXiv:cond-mat/1010.4965.
13. Kurose, T. On the Divergence of 1-conformally Flat Statistical Manifolds. *Tôhoku Math. J.* **1994**, *46*, 427–433.
14. Matsuzoe, H. Geometry of contrast functions and conformal geometry. *Hiroshima Math. J.* **1999**, *29*, 175–191.
15. Kurose, T. Conformal-projective geometry of statistical manifolds. *Interdisciplinary Information Sciences* **2002**, *8*, 89–100.
16. Yamano, T. Information theory based on non-additive information content. *Phys. Rev. E* **2001**, *63*, 046105.
17. Naudts, J. Estimators, escort probabilities, and phi-exponential families in statistical physics. *J. Ineq. Pure Appl. Math.* **2004**, *5*, 102.
18. Pistone, G. kappa-exponential models from the geometrical viewpoint. *Eur. Phys. J. B* **2009**, *70*, 29–37.
19. Naudts, J. Generalized exponential families and associated entropy functions. *Entropy* **2008**, *10*, 131–149.
20. Kaniadakis, G.; Lissia, M.; Scarfone, A.M. Deformed logarithms and entropies. *Physica A* **2004**, *340*, 41–49.
21. Yamano, T. Some properties of  $q$ -logarithmic and  $q$ -exponential functions in Tsallis statistics. *Physica A* **2002**, *305*, 486–496.
22. Tsallis, C.; Levy, S.V.F.; Souza, A.M.C.; Maynard, R. Statistical-mechanical foundation of the ubiquity of Levy distributions in nature. *Phys. Rev. Lett.* **1995**, *75*, 3589–3593, Erratum *Phys. Rev. Lett.* **1996**, *77*, 5442.
23. Tanaka, M. A consideration on the family of  $q$ -Gaussian distributions. *IEICE (Japan)* **2002**, *J85-D2*, 161–173 (in Japanese).

24. Zhang, Z.; Zhong, F.; Sun, H. Information geometry of the power inverse Gaussian distribution. *Appl. Sci.* **2007**, *9*, 194–203.
25. Ohara, A.; Wada, T. Information geometry of  $q$ -Gaussian densities and behaviors of solutions to related diffusion equations. *J. Phys. A: Math. Theor.* **2010**, *43*, 035002.
26. Wada, T. Generalized  $\log$ -likelihood functions and Bregman divergences. *J. Math. Phys.* **2009**, *50*, 113301.
27. Cichocki, A.; Cruces, S.; Amari, S. Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization. *Entropy* **2011**, *13*, 134–170.
28. Amari, S.  $\alpha$ -divergence is unique, belonging to both  $f$ -divergence and Bregman divergence classes. *IEEE Trans. Inform. Theor.* **2009**, *55*, 4925–4931.
29. Beck, C.; Schlögl, F. *Thermodynamics of Chaotic Systems*; Cambridge University Press: Cambridge, UK, 1993.
30. Borland, L.; Plastino, A.R.; Tsallis C. Information gain within nonextensive thermostatics. *J. Math. Phys.* **1998**, *39*, 6490–6501.
31. Furuichi, S. Fundamental properties of Tsallis relative entropy. *J. Math. Phys.* **2004**, *45*, 4868–4877.
32. Okamoto, I.; Amari, S.; Takeuchi, K. Asymptotic theory of sequential estimation procedures for curved exponential families. *Ann. Stat.* **1991**, *19*, 961–981.
33. Rényi, A. On measures of entropy and information. In Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, Berkeley, CA, USA, 20 June–30 July 1960; pp. 547–561.
34. Landsberg, P.T.; Vedral, V. Distributions and channel capacities in generalized statistical mechanics. *Phys. Lett. A* **1998**, *247*, 211–217.
35. Rajagopal, A.K.; Abe, S. Implications of form invariance to the structure of nonextensive entropies. *Phys. Rev. Lett.* **1999**, *83*, 1711–1714.
36. Yamano, T. Source coding theorem based on a nonadditive information content. *Physica A* **2002**, *305*, 190–195.
37. Wada, T.; Scarfone, A.M. Connections between Tsallis' formalisms employing the standard linear average energy and ones employing the normalized  $q$ -average energy. *Phys. Lett. A* **2005**, *335*, 351–362.



### Conformal geometry of escort probabilities and its application to Voronoi partitions

A. Ohara<sup>†</sup>, H. Matsuzoe, S. Amari

<sup>†</sup>Osaka University

Escort probability is naturally induced from researches of multifractals [1] and nonextensive statistical mechanics [2] to play an important but mysterious role. Testing its utility in the other scientific fields would greatly help our understanding about it. This motivates us to approach the escort probability by geometrically studying its role in information science. The first purpose of this presentation is to investigate the escort probability from viewpoints of information geometry [3] and affine differential geometry [4]. The second is to show that escort probability with information geometric structure is useful to construction of Voronoi partitions (or diagrams) [5] on the space of probability distributions. Recently, it is reported [6] that alpha-geometry, which is an information geometric structure of constant curvature, has a close relation with Tsallis statistics [2]. The remarkable feature of the alpha-geometry consists of the Fisher metric together with a one parameter family of dual affine connections, called the alpha-connections. We prove that the manifold of escort probability distributions is dually flat by considering conformal transformations that flatten the alpha-geometry on the manifold of usual probability distributions. On the resultant manifold, escort probabilities consist of an affine coordinate system. The result gives us a clear geometrical interpretation of the escort probability, and simultaneously, produces its new obscure links to conformality and projectivity. Due to these two geometrical concepts, however, the obtained dually flat structure inherits several properties of the alpha-geometry. The dually flatness proves crucial to construction of Voronoi partitions for alpha-divergences, which we shall call alpha-Voronoi partitions. The Voronoi partitions on the space of probability distributions with the Kullback-Leibler, or Bregman divergences have been recognized as important tools for various statistical modeling problems involving pattern classification, clustering, likelihood ratio test and so on. The largest advantage to take account of alpha-divergences is their invariance under transformations by sufficient statistics (See also [3] in a different viewpoint), which is a significant requirement for those statistical applications. In computational aspect, the conformal flattening of the alpha-geometry enables us to invoke the standard algorithm [5] using a potential function and an upper envelop of hyperplanes with the escort probabilities as coordinates.

- [1] C. Beck and F. Schlogl, *Thermodynamics of Chaotic Systems*, (Cambridge University Press, 1993).
- [2] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, (Berlin/Heidelberg: Springer, 2009).
- [3] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*, (Rhode Island: AMS&Oxford, 2000).
- [4] K. Nomizu and T. Sasaki, *Affine Differential Geometry*, (Cambridge University, 1993).
- [5] H. Edelsbrunner, *Algorithms in Combinatorial Geometry*, (Springer-Verlag, 1987).
- [6] A. Ohara A, Phys. Lett. A **370** 184 (2007); Euro. Phys. J. B **70** 15 (2009).

### Characteristics of bubble in house price distribution of Japan

T. Ohnishi<sup>†</sup>, T. Mizuno, C. Shimizu, T. Watanabe

<sup>†</sup>Canon Institute for Global Studies and University of Tokyo

We empirically investigate the house price distributions in the Greater Tokyo Area by using a housing information which is published on a weekly basis by Recruit Co., Ltd.. This dataset contains individual listings of 724,416 condominiums from 1986 to 2009 including the period of housing bubble. The attributes of a house are also included such as its size, location, age, and so on. This dataset covers more than 95 percent of the entire transactions in the central part of Tokyo (the 23 special wards of Tokyo). We find that the cross-sectional distribution of house prices has a fat upper tail, and the tail part is close to that of a power law distribution with exponent  $\approx 3$ . On the other hand, the cross-sectional distribution of house sizes measured in terms of floor space has less fat tails than the price distribution and is close to an exponential distribution with mean 25 square meters. We also find a positive linear relationship between the log price of a house and its size. An increase in the house size by a square meter leads to a 1.3 percent increase in the house price. We construct a size-adjusted price by subtracting the house size (multiplied by a positive coefficient) from the log price, which is consistent with these findings. We find that the size-adjusted price follows a lognormal distribution except for the period of the asset bubble and its collapse in Tokyo for which the price distribution remains asymmetric and skewed to the right even after controlling for the size effect. As for the period of the bubble and its collapse, we find some evidence that the sharp price movements were concentrated in particular areas, and this spatial heterogeneity is the source of the fat upper tail. These findings show that the cross-sectional distribution of size-adjusted prices is very close to a lognormal distribution during regular times but deviated substantially from a lognormal during the bubble period. This suggests that the shape of the size-adjusted price distribution, especially the shape of the tail part, may contain information useful for the detection of housing bubbles. That is, the presence of a bubble can be safely ruled out if recent price observations are found to follow a lognormal distribution. On the other hand, if there are many outliers, especially near the upper tail, this may indicate the presence of a bubble, since such price observations are very unlikely to occur if they follow a lognormal distribution. This method of identifying bubbles is quite different from conventional ones based on aggregate measures of housing prices, and therefore should be a useful tool to supplement existing methods.

- [1] T. Ohnishi, T. Mizuno, C. Shimizu and T. Watanabe, *On the evolution of the house price distribution*, Research Center for Price Dynamics (Hitotsubashi University), Discussion Paper No. 56 (2010).
- [2] C.-I. Chou and S.P. Li, *House price distributions of Taiwan: a preliminary study*, arXiv:1008.1376v1
- [3] N. MacKay, *London house prices are power-law distributed*, arXiv:1012.3039v1

# Dually flat structure with escort probability and its application to alpha-Voronoi diagrams<sup>‡</sup>

Atsumi Ohara<sup>1</sup>, Hiroshi Matsuzoe<sup>2</sup> and Shun-ichi Amari<sup>3</sup>

<sup>1</sup>Department of Systems Science, Osaka University, Toyonaka, Osaka 560-8531, Japan

E-mail: ohara@sys.es.osaka-u.ac.jp

<sup>2</sup>Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, 466-8555 Japan

E-mail: matsuzoe@nitech.ac.jp

<sup>3</sup>Riken Brain Science Institute, Wako-shi Hirosawa 2-1, Saitama 351-0198 Japan

E-mail: amari@brain.riken.jp

**Abstract.** This paper studies geometrical structure of the manifold of escort probability distributions and shows its new applicability to information science. In order to realize escort probabilities we use a conformal transformation that flattens so-called alpha-geometry of the space of discrete probability distributions, which well characterizes nonadditive statistics on the space. As a result escort probabilities are proved to be flat coordinates of the usual probabilities for the derived dually flat structure. Finally, we demonstrate that escort probabilities with the new structure admits a simple algorithm to compute Voronoi diagrams and centroids with respect to alpha-divergences.

PACS numbers: 05.90.+m, 89.70.Cf, 02.40.Hw

<sup>‡</sup> Several results in this paper can be found in the conference paper [36] without complete proofs.

## 1. Introduction

Escort probability is naturally induced from researches of multifractals [1] and non-extensive statistical mechanics [2] to play an important but mysterious role. Testing its utility in the other scientific fields would greatly help our understanding about it. This motivates us to approach the escort probability by geometrically studying its role in information science.

The first purpose of this paper is to investigate the escort probability from viewpoints of information geometry [3, 4] and affine differential geometry [5]. The second is to show that escort probability with information geometric structure is useful to construction of Voronoi diagrams [6] on the space of probability distributions.

Recently, it is reported [7, 8] that  $\alpha$ -geometry, which is an information geometric structure of constant curvature, has a close relation with Tsallis statistics [2]. The remarkable feature of the  $\alpha$ -geometry consists of the Fisher metric together with a one-parameter family of dual affine connections, called the  $\alpha$ -connections.

We prove that the manifold of escort probability distributions is dually flat by considering conformal transformations that flatten the  $\alpha$ -geometry on the manifold of usual probability distributions. On the resultant manifold, escort probabilities consist of an affine coordinate system. See also [9] for another type of flattening a curved dual manifold by a conformal transformation.

The result gives us a clear geometrical interpretation of the escort probability, and simultaneously, produces its new obscure links to conformality and projectivity. Due to these two geometrical concepts, however, the obtained dually flat structure inherits several properties of the  $\alpha$ -geometry.

The dually flatness proves crucial to construction of Voronoi diagrams for  $\alpha$ -divergences, which we shall call  $\alpha$ -Voronoi diagrams. The Voronoi diagrams on the space of probability distributions with the Kullback-Leibler [10, 11], or Bregman divergences [12] have been recognized as important tools for various statistical modeling problems involving pattern classification, clustering, likelihood ratio test and so on. See also, e.g., [13, 14, 15] for related problems.

The largest advantage to take account of  $\alpha$ -divergences is their *invariance under transformations by sufficient statistics* [16] (See also [4] in a different viewpoint), which is a significant requirement for those statistical applications. In computational aspect, the conformal flattening of the  $\alpha$ -geometry enables us to invoke the standard algorithm [29, 6] using a potential function and an upper envelop of hyperplanes with the escort probabilities as coordinates.

Section 2 is devoted to preliminaries for  $\alpha$ -geometry in the light of affine differential geometry. In section 3, as a main result, we consider conformal transformations and discuss properties of the obtained dually flat structure. Dual pairs of potential functions and affine coordinate systems on the manifold are explicitly identified, and the associated canonical divergence is shown to be conformal to the  $\alpha$ -divergence. Section 4 describes an application of such a flattened geometric structure to  $\alpha$ -Voronoi diagrams on the

probability simplex. The properties and a construction algorithm are discussed. Further, a formula for  $\alpha$ -centroid is touched upon.

In the sequel, we fix the relations of two parameters  $q$  and  $\alpha$  as  $q = (1 - \alpha)/2$ , and restrict  $q > 0$ .

## 2. Preliminaries

We briefly introduce  $\alpha$ -geometry via affine differential geometry. See for details [7, 8]. Let  $\mathcal{S}^n$  denote the  $n$ -dimensional probability simplex, i.e.,

$$\mathcal{S}^n := \left\{ \mathbf{p} = (p_i) \left| p_i > 0, \sum_{i=1}^{n+1} p_i = 1 \right. \right\}, \quad (1)$$

and  $p_i, i = 1, \dots, n+1$  denote probabilities of  $n+1$  states. We introduce the  $\alpha$ -geometric structure on  $\mathcal{S}^n$ . Let  $\{\partial_i\}, i = 1, \dots, n$  be natural basis tangent vector fields on  $\mathcal{S}^n$  defined by

$$\partial_i := \frac{\partial}{\partial p_i} - \frac{\partial}{\partial p_{n+1}}, \quad i = 1, \dots, n, \quad (2)$$

where  $p_{n+1} = 1 - \sum_{i=1}^n p_i$ . Now we define a Riemannian metric  $g$  on  $\mathcal{S}^n$  called the *Fisher metric*:

$$\begin{aligned} g_{ij}(\mathbf{p}) &:= g(\partial_i, \partial_j) = \frac{1}{p_i} \delta_{ij} + \frac{1}{p_{n+1}} \\ &= \sum_{k=1}^{n+1} p_k (\partial_i \log p_k) (\partial_j \log p_k), \quad i, j = 1, \dots, n. \end{aligned} \quad (3)$$

Further, define an torsion-free affine connection  $\nabla^{(\alpha)}$  called the  $\alpha$ -connection, which is represented in its coefficients by

$$\Gamma_{ij}^{(\alpha)k}(\mathbf{p}) = \frac{1+\alpha}{2} \left( -\frac{1}{p_k} \delta_{ij}^k + p_k g_{ij} \right), \quad i, j, k = 1, \dots, n, \quad (4)$$

where  $\delta_{ij}^k$  is equal to one if  $i = j = k$  and zero otherwise. Then we have the  $\alpha$ -covariant derivative  $\nabla^{(\alpha)}$ , which gives

$$\nabla_{\partial_i}^{(\alpha)} \partial_j = \sum_{k=1}^n \Gamma_{ij}^{(\alpha)k} \partial_k,$$

when it is applied to the vector fields  $\partial_i$  and  $\partial_j$ .

There are two specific features for the  $\alpha$ -geometry on  $\mathcal{S}^n$  defined in such a way. First, the triple  $(\mathcal{S}^n, g, \nabla^{(\alpha)})$  is a *statistical manifold* [17] (See appendix A for its definition), i.e., we can confirm that the following relation holds:

$$Xg(Y, Z) = g(\nabla_X^{(\alpha)} Y, Z) + g(Y, \nabla_X^{(-\alpha)} Z), \quad X, Y, Z \in \mathcal{X}(\mathcal{S}^n), \quad (5)$$

where  $\mathcal{X}(\mathcal{S}^n)$  denotes the set of all tangent vector fields on  $\mathcal{S}^n$ . Two statistical manifolds  $(\mathcal{S}^n, g, \nabla^{(\alpha)})$  and  $(\mathcal{S}^n, g, \nabla^{(-\alpha)})$  are said *mutually dual*.

The other is that  $(\mathcal{S}^n, g, \nabla^{(\alpha)})$  is a manifold of constant curvature  $\kappa = (1 - \alpha^2)/4$ , i.e.,

$$R^{(\alpha)}(X, Y)Z = \kappa \{g(Y, Z)X - g(X, Z)Y\},$$



where  $R^{(\alpha)}$  is the curvature tensor with respect to  $\nabla^{(\alpha)}$ . From this property the well-known nonadditive formula of the Tsallis entropy can be derived [7].

In [8] we have discussed the  $\alpha$ -geometry on  $\mathcal{S}^n$  from a viewpoint of *affine differential geometry* [5]. Consider the immersion  $f$  of  $\mathcal{S}^n$  into  $\mathbf{R}_+^{n+1}$  by

$$f : \mathbf{p} = (p_i) \mapsto \mathbf{x} = (x^i) = (L^{(\alpha)}(p_i)), \quad i = 1, \dots, n+1, \quad (6)$$

where  $(x^i), i = 1, \dots, n+1$  is the canonical flat coordinate system of  $\mathbf{R}^{n+1}$  and the function  $L^{(\alpha)}$  is defined by

$$L^{(\alpha)}(t) := \frac{2}{1-\alpha} t^{(1-\alpha)/2} = \frac{1}{q} t^q.$$

Note that  $f(\mathcal{S}^n)$  is a level hypersurface in the ambient space  $\mathbf{R}_+^{n+1}$  represented by  $\Psi(\mathbf{x}) = 2/(1+\alpha)$ , where

$$\Psi(\mathbf{x}) := \frac{2}{\alpha+1} \sum_{i=1}^{n+1} \left( \frac{1-\alpha}{2} x^i \right)^{2/(1-\alpha)} = \frac{1}{1-q} \sum_{i=1}^{n+1} (q x^i)^{1/q}. \quad (7)$$

We choose a transversal vector  $\xi$  on the level hypersurface by

$$\xi := \sum_{i=1}^{n+1} \xi^i \frac{\partial}{\partial x^i}, \quad \xi^i = -q(1-q)x^i = -\kappa x^i. \quad (8)$$

Then we can confirm that the affine immersion  $(f, \xi)$  realizes the  $\alpha$ -geometry on  $\mathcal{S}^n$  [8]. Hence, it would be possible to develop theory of the  $\alpha$ -geometry and Tsallis statistics with ideas of affine differential geometry [18].

Further, the *escort probability* [1] naturally appears in this setup. The escort probability  $\mathbf{P} = (P_i)$  associated with  $\mathbf{p} = (p_i)$  is the normalized version of  $(p_i)^q$ , and is defined by

$$P_i(\mathbf{p}) := \frac{(p_i)^q}{\sum_{j=1}^{n+1} (p_j)^q} = \frac{x^i}{Z_q}, \quad i = 1, \dots, n+1, \quad Z_q(\mathbf{p}) := \sum_{i=1}^{n+1} x^i(\mathbf{p}), \quad \mathbf{x}(\mathbf{p}) \in f(\mathcal{S}^n). \quad (9)$$

Hence, the simplex  $\mathcal{E}^n$  in the ambient space  $\mathbf{R}_+^{n+1}$ , i.e.,

$$\mathcal{E}^n := \left\{ \mathbf{x} = (x^i) \left| \sum_{i=1}^{n+1} x^i = 1, x^i > 0 \right. \right\}$$

represents the set of escort distributions  $\mathbf{P}$ .

Note that the element  $\mathbf{x}^* = (x_i^*)$  in the dual space of  $\mathbf{R}^{n+1}$  defined by

$$x_i^*(\mathbf{p}) := L^{(-\alpha)}(p_i) = \frac{1}{1-q} (p_i)^{1-q}, \quad i = 1, \dots, n+1,$$

meets

$$x_i^*(\mathbf{p}) = \frac{\partial \Psi}{\partial x^i}(\mathbf{x}(\mathbf{p})).$$

Hence, it satisfies [8]

$$-\sum_{i=1}^{n+1} \xi^i(\mathbf{p}) x_i^*(\mathbf{p}) = 1, \quad \sum_{i=1}^{n+1} x_i^*(\mathbf{p}) X^i = 0, \quad (10)$$

for an arbitrary vector  $X = \sum_{i=1}^{n+1} X^i \partial / \partial x^i$  at  $\mathbf{x}(\mathbf{p})$  tangent to  $f(\mathcal{S}^n)$ . Thus,  $-\mathbf{x}^*(\mathbf{p})$  can be interpreted as the *conormal map* [5].

### 3. A conformally and projectively flat geometric structure and escort probabilities

In this section we show a main result. For this purpose, we consider a conformal and projective transformation [19, 20, 21, 22] of the  $\alpha$ -geometry to introduce a dually flat one. This flattening of the  $\alpha$ -geometry conserves some of its properties. The escort probabilities  $(P_i)$  are found to represent one of mutually dual affine coordinate systems in the induced geometry. While the many functions or geometric quantities introduced in this section depend on the parameter  $\alpha$  or  $q$ , we omit them for the brevity.

Let us define a function  $\lambda$  on  $\mathcal{S}^n$  by

$$\lambda(\mathbf{p}) := \frac{1}{Z_q} = \frac{1}{\sum_{i=1}^{n+1} L^{(\alpha)}(p_i)},$$

which depends on  $\alpha$ . Then, from (9)  $\mathcal{E}^n$  is regarded as the image of  $\mathcal{S}^n$  for another immersion  $\tilde{f} := \lambda f$ , i.e.,

$$\tilde{f} : \mathcal{S}^n \ni (p_i) \mapsto (P_i) \in \mathcal{E}^n, \quad i = 1, \dots, n+1,$$

and  $(P_1, \dots, P_n)$  is interpreted as another coordinate system of  $\mathcal{S}^n$ . Note that the inverse mapping  $\tilde{f}^{-1}$  is well-defined by

$$\tilde{f}^{-1} : (P_i) \mapsto (p_i) = \left( \frac{(P_i)^{1/q}}{\sum_{j=1}^{n+1} (P_j)^{1/q}} \right), \quad i = 1, \dots, n+1.$$

It would be a natural way to introduce geometric structure on  $\mathcal{E}^n$  (and hence on  $\mathcal{S}^n$ ) via the affine immersion  $(\tilde{f}, \tilde{\xi})$  by taking a suitable transversal vector  $\tilde{\xi}$ , similarly to the case of the  $\alpha$ -geometry mentioned above. Since  $\mathcal{E}^n$  is a part of a hyperplane in  $\mathbf{R}^{n+1}$ , the canonical affine connection of  $\mathbf{R}^{n+1}$  induces a flat connection, denoted by  $D^{(E)}$ , on  $\mathcal{E}^n$ . However, for the same reason, we cannot define a Riemannian metric in this way§ because it vanishes on  $\mathcal{E}^n$ , regardless of any choice of the transversal vector  $\tilde{\xi}$ .

The idea we adopt here is to define a Riemannian metric by utilizing a property of  $(\mathcal{S}^n, g, \nabla^{(\alpha)})$  called *-1-conformal flatness*. Based on the results proved by Kurose [19, 20], we conclude that the manifold  $(\mathcal{S}^n, g, \nabla^{(\alpha)})$  is  $\pm 1$ -conformally flat (See Appendix A for its definition) because it is a statistical manifold of constant curvature.

Actually, let  $\nabla^*$  be the flat connection|| on  $\mathcal{S}^n$  defined with  $D^{(E)}$  and the differential  $\tilde{f}_*$  by

$$\tilde{f}_*(\nabla_X^* Y) = D_{\tilde{f}_* X}^{(E)} \tilde{f}_* Y, \quad X, Y \in \mathcal{X}(\mathcal{S}^n).$$

Then, we can prove that  $\nabla^{(\alpha)}$  and  $\nabla^*$  are *projectively equivalent* [5], i.e., it holds that

$$\nabla_X^* Y = \nabla_X^{(\alpha)} Y + d(\ln \lambda)(Y)X + d(\ln \lambda)(X)Y, \quad X, Y \in \mathcal{X}(\mathcal{S}^n). \quad (11)$$

Hence, if we define another Riemannian metric  $h$  on  $\mathcal{S}^n$  by

$$h(X, Y) := \lambda g(X, Y), \quad X, Y \in \mathcal{X}(\mathcal{S}^n), \quad (12)$$

§ In affine differential geometry, a Riemannian metric is realized as the affine fundamental form of an affine immersion [5].

|| For the sake of notational consistency with the existing literature, e.g., [3, 4], we first define  $\nabla^*$ , and later  $\nabla$  as the dual of  $\nabla^*$ .

then,  $(\mathcal{S}^n, g, \nabla^{(\alpha)})$  is  $-1$ -conformally equivalent to  $(\mathcal{S}^n, h, \nabla^*)$  equipped with a flat connection  $\nabla^*$ . Further, the manifold  $(\mathcal{S}^n, h, \nabla^*)$  can be proved to be a statistical manifold (See Appendix B).

Using the conormal map  $-\mathbf{x}^*(\mathbf{p})$ , we can define the  $\alpha$ -divergence as a *contrast function* (See Appendix A) inducing  $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$  as follows [20]:

$$\begin{aligned} D^{(\alpha)}(\mathbf{p}, \mathbf{r}) &= - \sum_{i=1}^{n+1} x_i^*(\mathbf{r})(x^i(\mathbf{p}) - x^i(\mathbf{r})) \\ &= \langle -\mathbf{x}^*(\mathbf{r}), \mathbf{x}(\mathbf{p}) - \mathbf{x}(\mathbf{r}) \rangle = \frac{1}{\kappa} - \langle \mathbf{x}^*(\mathbf{r}), \mathbf{x}(\mathbf{p}) \rangle. \end{aligned}$$

The statistical manifolds  $(\mathcal{S}^n, g, \nabla^{(-\alpha)})$  and  $(\mathcal{S}^n, g, \nabla^{(\alpha)})$  are dual in the sense of (5). Further, it is known [4] that there exists the unique affine flat connection  $\nabla$  on  $\mathcal{S}^n$ , dual with respect to  $(h, \nabla^*)$ . Then, according to [20], it is proved that  $(\mathcal{S}^n, h, \nabla)$  is 1-conformally equivalent to  $(\mathcal{S}^n, g, \nabla^{(-\alpha)})$  and a contrast function  $\rho$  inducing  $(h, \nabla, \nabla^*)$  is given by scaling  $D^{(-\alpha)}$  (See Appendix A) as follows:

$$\begin{aligned} \rho(\mathbf{p}, \mathbf{r}) &= \lambda(\mathbf{r}) D^{(-\alpha)}(\mathbf{p}, \mathbf{r}) = \frac{1}{Z_q(\mathbf{r})} D^{(-\alpha)}(\mathbf{p}, \mathbf{r}) \\ &= \frac{1}{Z_q(\mathbf{r})} \langle -\mathbf{x}(\mathbf{r}), \mathbf{x}^*(\mathbf{p}) - \mathbf{x}^*(\mathbf{r}) \rangle = \langle -\mathbf{P}(\mathbf{r}), \mathbf{x}^*(\mathbf{p}) - \mathbf{x}^*(\mathbf{r}) \rangle. \end{aligned} \quad (13)$$

We shall call  $\rho$  a *conformal divergence*.

Now, since  $(\mathcal{S}^n, h, \nabla, \nabla^*)$  is a dually flat space, the standard result in [3, 4] suggests that there exist *mutually dual affine coordinate systems*  $(\theta^1, \dots, \theta^n)$  and  $(\eta_1, \dots, \eta_n)$ , a *potential function*  $\psi(\boldsymbol{\theta})$  and its *conjugate*  $\psi^*(\boldsymbol{\eta})$  satisfying

$$\eta_i = \frac{\partial \psi}{\partial \theta^i}, \quad \theta^i = \frac{\partial \psi^*}{\partial \eta_i}, \quad i = 1, \dots, n. \quad (14)$$

They completely determine dually flat structure, i.e., the coefficients of  $h$ ,  $\nabla$  and  $\nabla^*$  are derived as the second and third derivatives of  $\psi$  or  $\psi^*$ , for example,

$$\begin{aligned} h_{ij} &= h \left( \frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \theta^j} \right) = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}, \quad h^{ij} = h \left( \frac{\partial}{\partial \eta_i}, \frac{\partial}{\partial \eta_j} \right) = \frac{\partial^2 \psi^*}{\partial \eta_i \partial \eta_j}, \\ \Gamma_{ijk} &= h \left( \nabla_{\frac{\partial}{\partial \theta^i}} \frac{\partial}{\partial \theta^j}, \frac{\partial}{\partial \theta^k} \right) = 0, \quad \Gamma_{ijk}^* = h \left( \nabla_{\frac{\partial}{\partial \eta_i}}^* \frac{\partial}{\partial \eta_j}, \frac{\partial}{\partial \eta_k} \right) = \frac{\partial^3 \psi}{\partial \theta^i \partial \theta^j \partial \theta^k}, \end{aligned}$$

and so on. In order to identify  $\psi, \psi^*, \theta^i$  and  $\eta_i$  explicitly *without integrating*  $h_{ij}$  or  $h^{ij}$ , we shall search for them by examining whether the conformal divergence  $\rho$  can be represented in the form of the *canonical divergence* [4], i.e.,

$$\rho(\mathbf{p}, \mathbf{r}) = \psi(\boldsymbol{\theta}(\mathbf{p})) + \psi^*(\boldsymbol{\eta}(\mathbf{r})) - \sum_{i=1}^n \theta^i(\mathbf{p}) \eta_i(\mathbf{r}). \quad (15)$$

with the constraints (14). If this is possible, we can directly prove from (A.4) and (A.5) that the obtained  $\psi, \psi^*, (\theta^1, \dots, \theta^n)$  and  $(\eta_1, \dots, \eta_n)$  are pairs of dual potential functions and affine coordinate systems associated with  $(\mathcal{S}^n, h, \nabla, \nabla^*)$ .

Before showing the result, we define, for  $0 < q$  with  $q \neq 1$ , two functions by

$$\ln_q(s) := \frac{s^{1-q} - 1}{1 - q}, \quad s \geq 0, \quad \exp_q(t) := [1 + (1 - q)t]_+^{1/(1-q)}, \quad t \in \mathbf{R},$$

where  $[t]_+ := \max\{0, t\}$ , and the so-called Tsallis entropy [23] by

$$S_q(\mathbf{p}) := \frac{\sum_{i=1}^{n+1} (p_i)^q - 1}{1 - q}.$$

Note that  $s = \exp_q(\ln_q(s))$  holds and they respectively recover the usual logarithmic, exponential function and the Boltzmann-Gibbs-Shannon entropy  $-\sum_{i=1}^{n+1} p_i \ln p_i$  when  $q \rightarrow 1$ . For  $q > 0$ ,  $\ln_q(s)$  is concave on  $s > 0$ .

**Theorem 1** *For the dually flat space  $(\mathcal{S}^n, h, \nabla, \nabla^*)$  defined via  $\pm 1$ -conformal transformation from  $(\mathcal{S}^n, g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ , the associated potential functions  $\psi, \psi^*$ , and dually flat affine coordinate systems  $(\theta^1, \dots, \theta^n)$  and  $(\eta_1, \dots, \eta_n)$  are represented as follows:*

$$\begin{aligned} \theta^i(\mathbf{p}) &= x_i^*(\mathbf{p}) - x_{n+1}^*(\mathbf{p}), \quad i = 1, \dots, n \\ \eta_i(\mathbf{p}) &= P_i(\mathbf{p}), \quad i = 1, \dots, n \\ \psi(\boldsymbol{\theta}(\mathbf{p})) &= -\ln_q(p_{n+1}), \\ \psi^*(\boldsymbol{\eta}(\mathbf{p})) &= \frac{1}{\kappa} (\lambda(\mathbf{p}) - q) = \frac{1}{1 - q} \left( \sum_{i=1}^{n+1} (\eta_i)^{1/q} \right)^q - \frac{1}{1 - q}, \end{aligned}$$

where  $\kappa = (1 - \alpha^2)/4 = q(1 - q)$  is the scalar curvature of  $(\mathcal{S}^n, g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$  and  $\eta_{n+1} := P_{n+1}(\mathbf{p}) = 1 - \sum_{i=1}^n P_i(\mathbf{p})$ . Further, the coordinate systems  $(\theta^1, \dots, \theta^n)$  and  $(\eta_1, \dots, \eta_n)$  are  $\nabla$ - and  $\nabla^*$ -affine, respectively.

Proof) As is mentioned above we have only to check that the potential functions  $\psi, \psi^*$  and dual affine coordinates  $\theta^i, \eta_i$  in the statement satisfy (14) and (15) for the conformal divergence  $\rho$ . First, substitute them directly to the right-hand side of (15) and modify it caring for the relation  $\eta_{n+1} = 1 - \sum_{i=1}^n \eta_i$ , then we see that it coincides with  $\rho(\mathbf{p}, \mathbf{r})$  in (13). Next, since it holds that  $\ln_q(p_i) = x_i^*(\mathbf{p}) - 1/(1 - q)$ , we can alternatively represent

$$\theta^i(\mathbf{p}) = \ln_q(p_i) - \ln_q(p_{n+1}) = \ln_q(p_i) + \psi(\boldsymbol{\theta}(\mathbf{p})), \quad i = 1, \dots, n.$$

Hence, for  $\theta^{n+1} \equiv 0$  it holds

$$1 = \sum_{i=1}^{n+1} p_i = \sum_{i=1}^{n+1} \exp_q(\theta^i - \psi).$$

Differentiating the both sides by  $\theta^j, j = 1, \dots, n$ , we have

$$0 = \sum_{i=1}^{n+1} \left( \delta_{ij} - \frac{\partial \psi}{\partial \theta^j} \right) (p_i)^q = (p_j)^q - \frac{\partial \psi}{\partial \theta^j} \sum_{i=1}^{n+1} (p_i)^q, \quad j = 1, \dots, n.$$

Thus, the left equation of (14) holds. Finally, note that the conformal factor is represented by

$$\lambda(\mathbf{p}) = \frac{1}{Z_q(\mathbf{p})} = \frac{q}{\sum_{i=1}^{n+1} (p_i)^q} = \frac{q}{(\exp_q(S_q(\mathbf{p})))^{1-q}}. \quad (16)$$

Using the formula [24]:

$$\exp_q(S_q(\mathbf{p})) = \exp_{\frac{1}{q}} \left( S_{\frac{1}{q}}(\mathbf{P}) \right),$$

we see that

$$\lambda(\mathbf{p}) = q \left( \exp_{\frac{1}{q}} \left( S_{\frac{1}{q}}(\mathbf{p}) \right) \right)^{q-1} = q \left( \sum_{i=1}^{n+1} (P_i)^{\frac{1}{q}} \right)^q.$$

Hence, the second equality in the expression of  $\psi^*$  holds. The right equation of (14) follows if you again recall  $\eta_{n+1} = 1 - \sum_{i=1}^n \eta_i$ . Q.E.D.

**Corollary 1** *The escort probabilities  $P_i, i = 1, \dots, n$  are canonical affine coordinates of the flat affine connection  $\nabla^*$  on  $\mathcal{S}^n$ .*

*Remark 1:* Since the conformal factor  $\lambda$  in (16) can be alternatively represented by

$$\lambda(\mathbf{p}) = \frac{q}{(\exp_q(S_q(\mathbf{p})))^{1-q}} = \kappa \ln_q \left( \frac{1}{\exp_q(S_q(\mathbf{p}))} \right) + q,$$

we have another expression of  $\psi^*$ , i.e.,

$$\psi^* = \ln_q \left( \frac{1}{\exp_q(S_q(\mathbf{p}))} \right).$$

Thus, the potentials and dual coordinates given in the proposition recover the standard ones [3, 4] when  $q \rightarrow 1$ , i.e.,

$$\psi \rightarrow -\ln p_{n+1}, \quad \psi^* \rightarrow \sum_{i=1}^{n+1} p_i \log p_i \quad \theta^i \rightarrow \log(p_i/p_{n+1}), \quad \eta_i \rightarrow p_i, \quad i = 1, \dots, n.$$

Note that  $-\psi^*$  coincides with the entropy studied in [25, 26, 27] and referred to as the *normalized Tsallis entropy*. The conformal (or scaling) factor  $\lambda$  often appears in the study of the  $q$ -analysis.

*Remark 2:* Similarly to the above conformal transformation of  $(\mathcal{S}^n, g, \nabla^{(\alpha)})$ , we can define another one for  $(\mathcal{S}^n, g, \nabla^{(-\alpha)})$  with a conformal factor

$$\lambda'(\mathbf{p}) := \frac{1}{\sum_{i=1}^{n+1} L^{(-\alpha)}(p_i)},$$

and construct another dually flat structure  $(h' = \lambda'g, \nabla', \nabla'^*)$ . Hence, the following relations among them hold (See Figure 1).

	$(\mathcal{S}^n, h', \nabla')$	$\xleftrightarrow{\text{dual}}$	$(\mathcal{S}^n, h', \nabla'^*)$	
1-conformally equivalent	$\updownarrow$		$\updownarrow$	-1-conformally equivalent
	$(\mathcal{S}^n, g, \nabla^{(\alpha)})$	$\xleftrightarrow{\text{dual}}$	$(\mathcal{S}^n, g, \nabla^{(-\alpha)})$	
-1-conformally equivalent	$\updownarrow$		$\updownarrow$	1-conformally equivalent
	$(\mathcal{S}^n, h, \nabla^*)$	$\xleftrightarrow{\text{dual}}$	$(\mathcal{S}^n, h, \nabla)$	

**Figure 1.** Relations among geometries

*Remark 3:* Because of the projective equivalence (11), a submanifold in  $\mathcal{S}^n$  is  $\nabla^{(\alpha)}$ -autoparallel if and only if it is  $\nabla^*$ -autoparallel. In particular, the set of distributions constrained with the normalized  $q$ -expectations (escort averages) [2] is a simultaneously  $\nabla^{(\alpha)}$ - and  $\nabla^*$ -autoparallel submanifold in  $\mathcal{S}^n$ .

#### 4. Applications to construction of alpha-Voronoi diagrams and alpha-centroids

For given  $m$  points  $\mathbf{p}_1, \dots, \mathbf{p}_m$  on  $\mathcal{S}^n$  we define  $\alpha$ -Voronoi regions on  $\mathcal{S}^n$  using the  $\alpha$ -divergence as follows:

$$\text{Vor}^{(\alpha)}(\mathbf{p}_k) := \bigcap_{l \neq k} \{\mathbf{p} \in \mathcal{S}^n \mid D^{(\alpha)}(\mathbf{p}, \mathbf{p}_k) < D^{(\alpha)}(\mathbf{p}, \mathbf{p}_l)\}, \quad k = 1, \dots, m.$$

An  $\alpha$ -Voronoi diagram on  $\mathcal{S}^n$  is a collection of the  $\alpha$ -Voronoi regions and their boundaries. Note that  $D^{(\alpha)}$  approaches the Kullback-Leibler divergence if  $\alpha \rightarrow -1$ , and  $D^{(0)}$  is called the Hellinger distance. If we use the Rényi divergence of order  $\alpha \neq 1$  [28] defined by

$$D_\alpha(\mathbf{p}, \mathbf{r}) := \frac{1}{\alpha - 1} \ln \sum_{i=1}^{n+1} (p_i)^\alpha (r_i)^{1-\alpha},$$

instead of the  $\alpha$ -divergence,  $\text{Vor}^{(1-2\alpha)}(\mathbf{p}_k)$  gives the corresponding Voronoi region because of their one-to-one functional relationship.

The standard algorithm using projection of a polyhedron [29, 6] commonly works well to construct Voronoi diagrams for the Euclidean distance [6], the Kullback-Leibler [11] and Bregman divergences [12], respectively. The algorithm is applicable if a distance function is represented by the remainder of the first order Taylor expansion of a convex potential function in a suitable coordinate system. Geometrically speaking, this is satisfied if i) the divergence is a canonical one for a certain dually flat structure and ii) its affine coordinate system is chosen to realize the corresponding Voronoi diagrams. In this coordinate system with one extra complementary coordinate the polyhedron is expressed as the upper envelop of  $m$  hyperplanes tangent to the potential function.

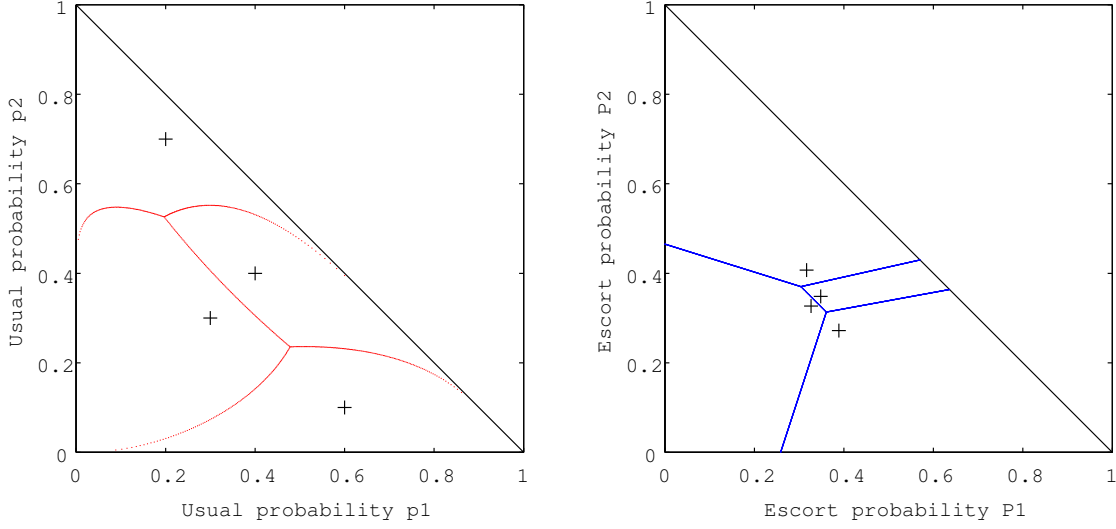
A problem for the case of the  $\alpha$ -Voronoi diagram is that the  $\alpha$ -divergence on  $\mathcal{S}^n$  cannot be represented as a remainder of any convex potentials. The following theorem, however, claims that the problem is resolved by conformally transforming the  $\alpha$ -geometry to the dually flat structure  $(h, \nabla, \nabla^*)$  and using the conformal divergence  $\rho$  and escort probabilities as a coordinate system.

Here, we denote the point on  $\mathcal{E}^n$  by  $\mathbf{P} = (P_1, \dots, P_n)$  because  $P_{n+1} = 1 - \sum_{i=1}^n P_i$ .

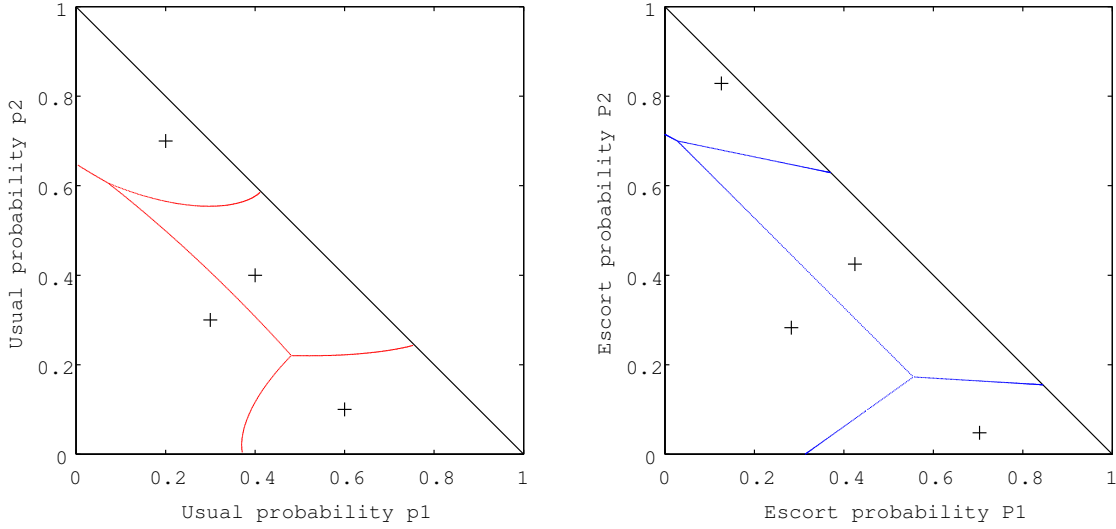
**Theorem 2** i) The bisector of  $\mathbf{p}_k$  and  $\mathbf{p}_l$  defined by  $\{\mathbf{p} \mid D^{(\alpha)}(\mathbf{p}, \mathbf{p}_k) = D^{(\alpha)}(\mathbf{p}, \mathbf{p}_l)\}$  is a simultaneously  $\nabla^{(\alpha)}$ - and  $\nabla^*$ -autoparallel hypersurface on  $\mathcal{S}^n$ .

ii) Let  $\mathcal{H}_k, k = 1, \dots, m$  be the hyperplane in  $\mathcal{E}^n \times \mathbf{R}$  which is respectively tangent at  $(\mathbf{P}_k, \psi^*(\mathbf{P}_k))$  to the hypersurface  $\{(\mathbf{P}, y) \mid y = \psi^*(\mathbf{P})\}$ , where  $\mathbf{P}_k = \mathbf{P}(\mathbf{p}_k)$ . The  $\alpha$ -Voronoi diagram can be constructed on  $\mathcal{E}^n$  as the projection of the upper envelope of  $\mathcal{H}_k$ 's along the  $y$ -axis.

Proof) i) Consider the  $\nabla^{(-\alpha)}$ -geodesic  $\gamma^{(-\alpha)}$  connecting  $\mathbf{p}_k$  and  $\mathbf{p}_l$ , and let  $\bar{\mathbf{p}}$  be the midpoint on  $\gamma^{(-\alpha)}$  satisfying  $D^{(\alpha)}(\bar{\mathbf{p}}, \mathbf{p}_k) = D^{(\alpha)}(\bar{\mathbf{p}}, \mathbf{p}_l)$ . Denote by  $\mathcal{B}$  the  $\nabla^{(\alpha)}$ -autoparallel hypersurface that is orthogonal to  $\gamma^{(-\alpha)}$  and contains  $\bar{\mathbf{p}}$ . Then, for all



**Figure 2.** An example of  $\alpha$ -Voronoi diagram on  $\mathcal{S}^2$  (left) for  $\alpha = 0.6$  (or  $q = 0.2$ ) and the corresponding one on  $\mathcal{E}^2$  (right).



**Figure 3.** An example of  $\alpha$ -Voronoi diagram on  $\mathcal{S}^2$  (left) for  $\alpha = -2$  (or  $q = 1.5$ ) and the corresponding one on  $\mathcal{E}^2$  (right).

$\mathbf{r} \in \mathcal{B}$ , the modified Pythagorean theorem [20, 7] implies the following equality:

$$\begin{aligned} D^{(\alpha)}(\mathbf{r}, \mathbf{p}_k) &= D^{(\alpha)}(\mathbf{r}, \bar{\mathbf{p}}) + D^{(\alpha)}(\bar{\mathbf{p}}, \mathbf{p}_k) - \kappa D^{(\alpha)}(\mathbf{r}, \bar{\mathbf{p}}) D^{(\alpha)}(\bar{\mathbf{p}}, \mathbf{p}_k) \\ &= D^{(\alpha)}(\mathbf{r}, \bar{\mathbf{p}}) + D^{(\alpha)}(\bar{\mathbf{p}}, \mathbf{p}_l) - \kappa D^{(\alpha)}(\mathbf{r}, \bar{\mathbf{p}}) D^{(\alpha)}(\bar{\mathbf{p}}, \mathbf{p}_l) = D^{(\alpha)}(\mathbf{r}, \mathbf{p}_l). \end{aligned}$$

Hence,  $\mathcal{B}$  is a bisector of  $\mathbf{p}_k$  and  $\mathbf{p}_l$ . The projective equivalence ensures that  $\mathcal{B}$  is also  $\nabla^*$ -autoparallel.

ii) Recall the equality  $D^{(\alpha)}(\mathbf{p}, \mathbf{r}) = D^{(-\alpha)}(\mathbf{r}, \mathbf{p})$  and the conformal relation (13) between  $D^{(-\alpha)}$  and  $\rho$ , then we see that  $\text{Vor}^{(\alpha)}(\mathbf{p}_k) = \text{Vor}^{(\text{conf})}(\mathbf{p}_k)$  holds on  $\mathcal{S}^n$ , where

$$\text{Vor}^{(\text{conf})}(\mathbf{p}_k) := \bigcap_{l \neq k} \{\mathbf{p} \in \mathcal{S}^n \mid \rho(\mathbf{p}_k, \mathbf{p}) < \rho(\mathbf{p}_l, \mathbf{p})\}.$$

Theorem 1, relations (14) and (15) imply that  $\rho(\mathbf{p}_k, \mathbf{p})$  is represented with the coordinates  $(P_i)$  by

$$\rho(\mathbf{p}_k, \mathbf{p}) = \psi^*(\mathbf{P}) - \left( \psi^*(\mathbf{P}_k) + \sum_{i=1}^n \frac{\partial \psi^*}{\partial P_i}(\mathbf{P}_k)(P_i(\mathbf{p}) - P_i(\mathbf{p}_k)) \right),$$

where  $\mathbf{P} = \mathbf{P}(\mathbf{p})$ . Note that a point  $(\mathbf{P}, y_k(\mathbf{P}))$  in  $\mathcal{H}_k$  is expressed by

$$y_k(\mathbf{P}) := \psi^*(\mathbf{P}_k) + \sum_{i=1}^n \frac{\partial \psi^*}{\partial P_i}(\mathbf{P}_k)(P_i(\mathbf{p}) - P_i(\mathbf{p}_k)).$$

Hence, we have  $\rho(\mathbf{p}_k, \mathbf{p}) = \psi^*(\mathbf{P}) - y_k(\mathbf{P})$ . We see, for example, that the bisector on  $\mathcal{E}^n$  for  $\mathbf{p}_k$  and  $\mathbf{p}_l$  is represented as a projection of  $\mathcal{H}_k \cap \mathcal{H}_l$ . Thus, the statement follows. Q.E.D.

The figure 2 and 3 show examples of  $\alpha$ -Voronoi diagrams on the simplex of dimension 2. In these cases, the bisectors are simultaneously  $\nabla^{(\alpha)}$ - and  $\nabla^*$ -geodesics.

*Remark 4:* In [30] Voronoi diagrams for broader class of divergences (contrast functions) that are not necessarily associated with any convex potentials are studied from more general affine differential geometric points of views. The construction algorithm is also given there, which is applicable if the corresponding affine immersion is explicitly obtained.

On the other hand, the  $\alpha$ -divergence defined not only on  $\mathcal{S}^n$  but on the positive orthant  $\mathbf{R}_+^{n+1}$  can be represented as a remainder of the potential  $\Psi$  in (7) [3, 4, 8]. Hence, the  $\alpha$ -geometry on  $\mathbf{R}_+^{n+1}$  is dually flat. Using this property,  $\alpha$ -Voronoi diagrams on  $\mathbf{R}_+^{n+1}$  is discussed in [31].

While both of the above methods require computation of the polyhedrons in the space of dimension  $n + 2$ , the new one proposed in this paper does in the space of dimension  $n + 1$ . Since the optimal computational time of polyhedrons depends on the dimension  $d$  by  $O(m \log m + m^{\lfloor d/2 \rfloor})$  [32], the new one where  $d = n + 1$  is slightly better when  $n$  is even.

The next proposition is a simple and relevant application of escort probabilities. Define the  $\alpha$ -centroid  $\mathbf{c}^{(\alpha)}$  for given  $m$  points  $\mathbf{p}_1, \dots, \mathbf{p}_m$  on  $\mathcal{S}^n$  by the minimizer of the following problem:

$$\min_{\mathbf{p} \in \mathcal{S}^n} \sum_{k=1}^m D^{(\alpha)}(\mathbf{p}_k, \mathbf{p}).$$

**Proposition 1** *The  $\alpha$ -centroid  $\mathbf{c}^{(\alpha)}$  for given  $m$  points  $\mathbf{p}_1, \dots, \mathbf{p}_m$  on  $\mathcal{S}^n$  is represented in escort probabilities by the weighted average of conformal factors  $\lambda(\mathbf{p}_k) = 1/Z_q(\mathbf{p}_k)$ , i.e.,*

$$P_i(\mathbf{c}^{(\alpha)}) = \frac{1}{\sum_{k=1}^m Z_q(\mathbf{p}_k)} \sum_{k=1}^m Z_q(\mathbf{p}_k) P_i(\mathbf{p}_k), \quad i = 1, \dots, n + 1.$$

Proof) Let  $\theta^i = \theta^i(\mathbf{p})$ . Using (13), (15) and the relation  $D^{(\alpha)}(\mathbf{p}, \mathbf{r}) = D^{(-\alpha)}(\mathbf{r}, \mathbf{p})$ , we have

$$\sum_{k=1}^m D^{(\alpha)}(\mathbf{p}_k, \mathbf{p}) = \sum_{k=1}^m Z_q(\mathbf{p}_k) \rho(\mathbf{p}, \mathbf{p}_k) = \sum_{k=1}^m Z_q(\mathbf{p}_k) \{ \psi(\boldsymbol{\theta}) + \psi^*(\boldsymbol{\eta}(\mathbf{p}_k)) - \sum_{i=1}^n \theta^i \eta_i(\mathbf{p}_k) \}.$$



Then the optimality condition is

$$\frac{\partial}{\partial \theta^i} \sum_{k=1}^m D^{(\alpha)}(\mathbf{p}_k, \mathbf{p}) = \sum_{k=1}^m Z_q(\mathbf{p}_k)(\eta_i - \eta_i(\mathbf{p}_k)) = 0, \quad i = 1, \dots, n,$$

where  $\eta_i = \eta_i(\mathbf{p})$ . Thus, the statement follows from Theorem 1 for  $i = 1, \dots, n$ . For  $i = n + 1$  it follows from the fact that the sum of the weights is equal to one. Q.E.D.

## 5. Concluding remarks

We have considered  $\pm 1$ -conformal transformations of the  $\alpha$ -geometry and obtained dually flat structure  $(\mathcal{S}^n, h, \nabla, \nabla^*)$ . Further the potential functions and dually flat coordinate systems associated with the structure have been derived. We see that the escort probability naturally appears to play an important role.

From a viewpoint of contrast functions, the geometric structure compatible to the Kullback-Leibler divergence is  $(\mathcal{S}^n, g, \nabla^{(1)}, \nabla^{(-1)})$ , where  $g$  is the Fisher information and  $\nabla^{(\pm 1)}$  are respectively the *e-connection* and the *m-connection*. Similarly, the  $\alpha$ -divergence (or the Tsallis relative entropy), and the conformal divergence  $\rho$  in this note correspond to  $(\mathcal{S}^n, g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$  and  $(\mathcal{S}^n, h, \nabla, \nabla^*)$ , respectively. They are summarized in Figure 4.

KL divergence	α-divergence	conformal divergence
$(\mathcal{S}^n, g, \nabla^{(1)}, \nabla^{(-1)})$	$(\mathcal{S}^n, g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$	$(\mathcal{S}^n, h, \nabla, \nabla^*), (\mathcal{S}^n, h', \nabla', \nabla'^*)$
dually flat	constant curvature $\kappa$	dually flat

**Figure 4.** transformations of dualistic structures

The physical meaning or essence underlying these transformations would be interesting and significant, but is left unclear. (See recent publications [33, 34] for such research directions.)

Finally, we have shown a direct application of the conformal flattening to computation of  $\alpha$ -Voronoi diagrams and  $\alpha$ -centroids. Escort probabilities are found to work as a suitable coordinate system for the purpose.

## Acknowledgments

The first author would like to thank Prof. Tatsuaki Wada for helpful comments.

## Appendix A: Statistical manifold and $\alpha$ -conformally equivalence

For details of this appendix see [17, 19, 20, 21, 22]. For a torsion-free affine connection  $\nabla$  and a pseudo Riemannian metric  $g$  on a manifold  $\mathcal{M}$ , the triple  $(\mathcal{M}, g, \nabla)$  is called a *statistical manifold* if it admits another torsion-free connection  $\nabla^*$  satisfying

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z) \quad (\text{A.1})$$

for arbitrary  $X, Y$  and  $Z$  in  $\mathcal{X}(\mathcal{M})$ , where  $\mathcal{X}(\mathcal{M})$  is the set of all tangent vector fields on  $\mathcal{M}$ . It is known that  $(\mathcal{M}, g, \nabla)$  is a statistical manifold if and only if  $\nabla g$  is symmetric, i.e.,  $(\nabla_X g)(Y, Z)$  is symmetric with respect to  $X, Y$  and  $Z$ . We call  $\nabla$  and  $\nabla^*$  *duals* of each other with respect to  $g$ , and  $(\mathcal{M}, g, \nabla^*)$  is said the dual statistical manifold of  $(\mathcal{M}, g, \nabla)$ . The triple of a Riemannian metric and a pair of dual connections  $(g, \nabla, \nabla^*)$  satisfying (A.1) is called a *dualistic structure* on  $\mathcal{M}$ .

For  $\alpha \in \mathbf{R}$ , statistical manifolds  $(\mathcal{M}, g, \nabla)$  and  $(\mathcal{M}, g', \nabla')$  are said to be  $\alpha$ -conformally equivalent if there exists a positive function  $\phi$  on  $\mathcal{M}$  such that

$$\begin{aligned} g'(X, Y) &= \phi g(X, Y), \\ g(\nabla'_X Y, Z) &= g(\nabla_X Y, Z) - \frac{1+\alpha}{2} d(\ln \phi)(Z)g(X, Y) \\ &\quad + \frac{1-\alpha}{2} \{d(\ln \phi)(X)g(Y, Z) + d(\ln \phi)(Y)g(X, Z)\}. \end{aligned}$$

Statistical manifolds  $(\mathcal{M}, g, \nabla)$  and  $(\mathcal{M}, g', \nabla')$  are  $\alpha$ -conformally equivalent if and only if  $(\mathcal{M}, g, \nabla^*)$  and  $(\mathcal{M}, g, \nabla'^*)$  are  $-\alpha$ -conformally equivalent.

A statistical manifold  $(\mathcal{M}, g, \nabla)$  is called  $\alpha$ -conformally flat if it is locally  $\alpha$ -conformally equivalent to a flat statistical manifold. Note that  $-1$ -conformal equivalence implies projective equivalence. A statistical manifold of dimension greater than three has constant curvature if and only if it is  $\pm 1$ -conformally flat.

We call a function  $\rho$  on  $\mathcal{M} \times \mathcal{M}$  a *contrast function* [35] inducing  $(g, \nabla, \nabla^*)$  if it satisfies

$$\rho(p, p) = 0, \quad p \in \mathcal{M}, \quad (\text{A.2})$$

$$\rho[X] = \rho[Y] = 0, \quad (\text{A.3})$$

$$g(X, Y) = -\rho[X|Y], \quad (\text{A.4})$$

$$g(\nabla_X Y, Z) = -\rho[XY|Z], \quad g(Y, \nabla_X^* Z) = -\rho[Y|XZ], \quad (\text{A.5})$$

where

$$\rho[X_1 \cdots X_k | Y_1 \cdots Y_l](p) := (X_1)_p \cdots (X_k)_p (Y_1)_q \cdots (Y_l)_q \rho(p, q)|_{p=q}$$

for arbitrary  $p, q \in \mathcal{M}$  and  $X_i, Y_j \in \mathcal{X}(\mathcal{M})$ . If  $(\mathcal{M}, g, \nabla)$  and  $(\mathcal{M}, g', \nabla')$  are 1-conformally equivalent, a contrast function  $\rho'$  inducing  $(g', \nabla', \nabla'^*)$  is represented by  $\rho$  inducing  $(g, \nabla, \nabla^*)$ , as

$$\rho'(p, q) = \phi(q)\rho(p, q).$$

## Appendix B: The proof for the fact that $(\mathcal{S}^n, h, \nabla^*)$ is a statistical manifold

We show that  $\nabla^* h$  is symmetric. By the definition of  $-1$ -conformally flatness we have

$$\begin{aligned} (\nabla_X^* h)(Y, Z) &= Xh(Y, Z) - h(\nabla_X^* Y, Z) - h(Y, \nabla_X^* Y) \\ &= d\lambda(X)g(Y, Z) + \lambda Xg(Y, Z) \\ &\quad - \lambda \{g(\nabla_X^{(\alpha)} Y, Z) + d(\ln \lambda)(Y)g(X, Z) + d(\ln \lambda)(X)g(Y, Z)\} \\ &\quad - \lambda \{g(Y, \nabla_X^{(\alpha)} Z) + d(\ln \lambda)(Z)g(X, Y) + d(\ln \lambda)(X)g(Z, Y)\}. \end{aligned}$$

Substitute the equality  $\lambda d(\ln \lambda) = d\lambda$  into the right-hand side, then it is transformed to

$$\begin{aligned} & \lambda \{Xg(Y, Z) - g(\nabla_X^{(\alpha)} Y, Z) - g(Y, \nabla_X^{(\alpha)} Z) \\ & \quad - d(\ln \lambda)(X)g(Y, Z) - d(\ln \lambda)(Y)g(X, Z) - d(\ln \lambda)(Z)g(X, Y)\} \\ & = \lambda(\nabla_X^{(\alpha)} g)(Y, Z) - \lambda \{d(\ln \lambda)(X)g(Y, Z) + d(\ln \lambda)(Y)g(X, Z) + d(\ln \lambda)(Z)g(X, Y)\}. \end{aligned}$$

Thus,  $\nabla^* h$  is symmetric because  $(\mathcal{S}^n, g, \nabla^{(\alpha)})$  is a statistical manifold, i.e.,  $\nabla^{(\alpha)} g$  is symmetric. Since  $\nabla^{(\alpha)}$  is torsion-free, so is  $\nabla^*$  by the definition of  $-1$ -conformally flatness.

## References

- [1] Beck C and Schlögl F 1993 *Thermodynamics of Chaotic Systems* (Cambridge: Cambridge University Press)
- [2] Tsallis C 2009 *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World* (Berlin/Heidelberg: Springer)
- [3] Amari S-I 1985 *Differential-Geometrical Methods in Statistics*, (Berlin Springer-Verlag)
- [4] Amari S-I and Nagaoka H 2000 *Methods of Information Geometry*, (Rhode Island: AMS&Oxford).
- [5] Nomizu K and Sasaki T 1993 *Affine Differential Geometry* (Cambridge: Cambridge University Press)
- [6] Edelsbrunner H 1987 *Algorithms in Combinatorial Geometry* (Springer-Verlag)
- [7] Ohara A 2007 Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization *Phys. Lett. A* **370** 184–93
- [8] Ohara A 2009 Geometric study for the Legendre duality of generalized entropies and its application to the porous medium equation *Euro. Phys. J. B* **70** 15–28
- [9] Okamoto I, Amari S and Takeuchi K 1991 Asymptotic theory of sequential estimation procedures for curved exponential families *Annals of Statistics* **19** 961–81
- [10] Onishi K and Takayama N 1996 Construction of Voronoi diagram on the Upper Half-Plane *IEICE Trans. on Fundamentals* **E79-A** 533–9
- [11] Onishi K and Imai H 1997 Voronoi diagram in statistical parametric space by Kullback-Leibler divergence *Proc. 13th ACM-SIAM Symposium on Computational Geometry* 463–5
- [12] Nielsen F, Boissonnat J-D and Nock N 2007 On Bregman Voronoi diagram *Proc. 18th ACM-SIAM Symposium on Discrete Algorithms* 746–55
- [13] Imai H and Inaba M 2000 Divergence-Based Geometric Clustering and Its Underlying Discrete Proximity Structures *IEICE Trans. Inf. & Syst.* **E83-D** 27–35
- [14] Lebanon G 2005 Metric Learning for Text Documents *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **28**, 497–508
- [15] Amari S-I 2007 Integration of stochastic models by minimizing  $\alpha$ -divergence *Neural Comput.*, vol. 19, no. 10, 2780–2796
- [16] Čencov N N, 1982 *Statistical Decision Rules and Optimal Inference*, AMS, Rhode Island (Originally published in Russian Nauka, Moscow (1972)).
- [17] Lauritzen S L 1987 Statistical manifolds S.Amari *et al* eds. *Differential geometry in statistical inference* 163–216 (Hayward: Institute of Mathematical Statistics)
- [18] Matsuzoe H, Takeuchi J and Amari S-I 2006 Equiaffine structures on statistical manifolds and Bayesian statistics *Differential Geom. Appl.* **24** 567–78
- [19] Kurose T 1990 Dual Connections and Affine Geometry *Math. Z.* **203** 115–21
- [20] Kurose T 1994 On the Divergence of 1-conformally Flat Statistical Manifolds, *Tôhoku Math. J.* **46** 427–33
- [21] Kurose T 2002 Conformal-projective geometry of statistical manifolds. *Interdiscip. Inform. Sci.* **8** 89–100

- [22] Matsuzoe H 1999 Geometry of contrast functions and conformal geometry *Hiroshima Math. J.* **29** 175–91
- [23] Tsallis C 1988 Possible generalization of Boltzmann-Gibbs statistics *J. Stat. Phys.* **52**, 479–87
- [24] Suyari H and Wada T 2008 Multiplicative duality,  $q$ -triplet and  $(\mu, \nu, q)$ -relation derived from the one-to-one correspondence between the  $(\mu, \nu)$ -multinomial coefficient and Tsallis entropy  $S_q$  *Physica A* **387** 71–83
- [25] Landsberg P T and Vedral V 1998 Distributions and channel capacities in generalized statistical mechanics *Phys. Lett. A* **247** 211–7
- [26] Rajagopal A K and Abe S 1999 Implications of Form Invariance to the Structure of Nonextensive Entropies *Phys. Rev. Lett.* **83** 1711–4
- [27] Wada T and Scarfone A M 2005 Connections between Tsallis' formalisms employing the standard linear average energy and ones employing the normalized  $q$ -average energy *Phys. Lett. A* **335** 351–62
- [28] Rényi A 1961 On measures of entropy and information *Proc. 4th Berkeley Symp. Math. Statist. Probability* **1** 547–61 (Univ. Calif. Press: Berkeley)
- [29] Edelsbrunner H and Seidel R 1986 *Discrete Comput. Geom.* **1** 25–44
- [30] Matsuzoe H 2009 Computational Geometry from the Viewpoint of Affine Differential Geometry *F.Nielsen eds. Emerging Trends in Visual Computing* 103–23 (Berlin/Heidelberg: Springer)
- [31] Nielsen, F and Nock, R 2009 The dual Voronoi diagrams with respect to representational Bregman divergences. International Symposium on Voronoi Diagrams (ISVD), DTU Lyngby, Denmark, June 2009. IEEE Press, New York
- [32] Chazelle B, 1993 An optimal convex hull algorithm in any fixed dimension, *Discrete and Computational Geometry* **10** 377–409
- [33] Ojima I 2010 Meaning of Non-Extensive Entropies in Micro-Macro Duality *J. Phys.: Conf. Ser.* **201** 012017
- [34] Tanaka M 2010 Meaning of an escort distribution and  $\tau$ -transformation *J. Phys.: Conf. Ser.* **201** 012007
- [35] Eguchi S 1992 Geometry of minimum contrast *Hiroshima Math. J.* **23** 631–47
- [36] Ohara A, Matsuzoe H and Amari S 2010 A dually flat structure on the space of escort distributions *J. Phys.: Conf. Ser.* **201** 012012

## GEOMETRY FOR $q$ -EXPONENTIAL FAMILIES

Hiroshi MATSUZOE

*Department of Computer Science and Engineering,  
Graduate School of Engineering, Nagoya Institute of Technology,  
Nagoya, Aichi 466-8555 Japan  
E-mail: matsuzoe@nitech.ac.jp*

Atsumi OHARA

*Department of Electrical and Electronics Engineering,  
Graduate School of Engineering, University of Fukui,  
Fukui 910-8507 Japan  
E-mail: ohara@fuee.u-fukui.ac.jp*

Geometry for  $q$ -exponential families is studied in this paper. A  $q$ -exponential family is a set of probability distributions, which is a natural generalization of the standard exponential family. A  $q$ -exponential family has information geometric structure and a dually flat structure. To describe these relations, generalized conformal structures for statistical manifolds are studied in this paper. As an application of geometry for  $q$ -exponential families, a geometric generalization of statistical inference is also studied.

*Keywords:*  $q$ -exponential family,  $q$ -product, Information geometry, Tsallis statistics, Statistical manifold, Divergence.

### Introduction

An exponential family is a set of probability distributions such as a set of normal distributions, of Poisson distributions, or of gamma distributions, etc. Such probability distributions decay exponentially. However, in complex systems, probability distributions often have long tails, that is, probability distributions do not decay exponentially. The  $q$ -normal distribution which is frequently discussed in Tsallis nonextensive statistical mechanics [18] is a typical example of such probability distributions.

In this paper, we consider  $q$ -exponential families. A  $q$ -exponential family is a natural generalization of the standard exponential family, and which includes the set of  $q$ -normal distributions. From the viewpoint of information geometry, it is known that an exponential family has a dually flat structure (see [1]). We will see that  $q$ -exponential families naturally have dually flat

structures.

A  $q$ -exponential family also has information geometric structure, that is, a  $q$ -exponential family has the Fisher metric and  $\alpha$ -connections. Hence a  $q$ -exponential family has two kinds of statistical manifold structures. Thus, we consider relations of these structures using generalized conformal equivalence relations on statistical manifolds.

In the later part of this paper, we consider statistical inferences for  $q$ -exponential families. Generalizations of independence or likelihood functions have been introduced in machine learning theory [4] or in Tsallis statistics [16]. We show that dually flat structures on  $q$ -exponential families work naturally for such generalized statistical inferences.

## 1. Preliminaries

In this section, we review geometry of statistical models and related geometry (cf.[1, 15]). We assume that all objects are smooth throughout this paper. We also assume that the manifold is simply connected since we will discuss geometry of statistical models.

### 1.1. Statistical models

Let  $\mathcal{X}$  be a total sample space and let  $\Xi$  be an open domain of  $\mathbf{R}^n$ . We say that  $S$  is a *statistical model* or a *parametric model* on  $\mathcal{X}$  if  $S$  is a set of probability densities with parameter  $\xi \in \Xi$  such that

$$S = \left\{ p(x; \xi) \left| \int_{\mathcal{X}} p(x; \xi) dx = 1, p(x; \xi) > 0, \xi \in \Xi \subset \mathbf{R}^n \right. \right\}.$$

Under suitable conditions,  $S$  can be regarded as a manifold with a local coordinate system  $\{\xi^1, \dots, \xi^n\}$  (see [1]).

For a statistical model  $S$ , we define a function  $g_{ij}^F(\xi) : \Xi \rightarrow \mathbf{R}$  by the following formula:

$$\begin{aligned} g_{ij}^F(\xi) &:= \int_{\mathcal{X}} \left( \frac{\partial}{\partial \xi^i} \log p(x; \xi) \right) \left( \frac{\partial}{\partial \xi^j} \log p(x; \xi) \right) p(x; \xi) dx \\ &= E_{\xi}[\partial_i l_{\xi} \partial_j l_{\xi}]. \end{aligned}$$

Here, for simplicity, we used following notations:

$$\begin{aligned} E_{\xi}[f] &= \int_{\mathcal{X}} f(x) p(x; \xi) dx, & (\text{the expectation of } f(x) \text{ at } p(x; \xi)), \\ l_{\xi} &= l(x; \xi) = \log p(x; \xi), & (\text{the log likelihood of } p(x; \xi)), \\ \partial_i &= \frac{\partial}{\partial \xi^i}. \end{aligned}$$

We assume that  $g_{ij}^F(\xi)$  is finite for all  $i, j, \xi$ . Set a matrix  $g^F = (g_{ij}^F)$ , then we can check that  $g^F$  is symmetric and non-negative definite. We assume that  $g^F$  is positive definite. Then  $g^F$  is a Riemannian metric on  $S$ . We call  $g^F$  the *Fisher metric* on  $S$ .

For  $\alpha \in \mathbf{R}$ , we define the  $\alpha$ -connection  $\nabla^{(\alpha)}$  by the following formulas:

$$\Gamma_{ij,k}^{(\alpha)}(\xi) = E_\xi \left[ \left( \partial_i \partial_j l_\xi + \frac{1-\alpha}{2} \partial_i l_\xi \partial_j l_\xi \right) (\partial_k l_\xi) \right],$$

$$h(\nabla_{\partial_i}^{(\alpha)} \partial_j, \partial_k) = \Gamma_{ij,k}^{(\alpha)}.$$

We can check that  $\nabla^{(\alpha)}$  is torsion-free and  $\nabla^{(0)}$  is the Levi-Civita connection of the Fisher metric. It is known that  $\pm 1$ -connections are more important than the Levi-Civita connection in geometric theory of statistical inferences. We call  $\nabla^{(1)}$  the *exponential connection* and  $\nabla^{(-1)}$  the *mixture connection*.

For  $\alpha$ -connections, the following formula holds

$$X g^F(Y, Z) = g^F(\nabla_X^{(\alpha)} Y, Z) + g^F(Y, \nabla_X^{(-\alpha)} Z).$$

The connections  $\nabla^{(\alpha)}$  and  $\nabla^{(-\alpha)}$  are said to be *dual* (or *conjugate*) with respect to  $g^F$ . For arbitrary  $\alpha, \beta \in \mathbf{R}$ , the difference between the  $\alpha$ -connection and the  $\beta$ -connection is given by

$$\Gamma_{ij,k}^{(\beta)} = \Gamma_{ij,k}^{(\alpha)} + \frac{\alpha - \beta}{2} C_{ijk}^F,$$

where

$$C_{ijk}^F(\xi) = E_\xi[\partial_i l_\xi \partial_j l_\xi \partial_k l_\xi].$$

The  $(0, 3)$ -tensor field  $C^F$  determined by  $C_{ijk}^F$  is called a *cubic form*. The covariant derivative of the Fisher metric  $g^F$  satisfies  $(\nabla_X^{(\alpha)} g^F)(Y, Z) = \alpha C^F(X, Y, Z)$ .

We say that a statistical model  $S$  is an *exponential family* if

$$S = \left\{ p(x; \theta) \mid p(x; \theta) = \exp \left[ Z(x) + \sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) \right], \theta \in \Theta \subset \mathbf{R}^n \right\},$$

where  $\Theta$  is a parameter space,  $Z, F_1, \dots, F_n$  are random variables on  $\mathcal{X}$  and  $\psi$  is a function on  $\Theta$ . The coordinate system  $\{\theta^i\}$  is called the *natural parameters*.

**Proposition 1.1.** *For an exponential family  $S$ , the natural parameters  $\{\theta^i\}$  is an affine coordinate system with respect to  $\nabla^{(1)}$ , that is,  $\Gamma_{ij}^{(1)k} \equiv 0$  ( $i, j, k = 1, \dots, n$ ), and the 1-connection  $\nabla^{(1)}$  is flat.*

For simplicity, we set  $Z = 0$ . It is possible to assume this condition without loss of generality. We say that  $M$  is a *curved exponential family* of  $S$  if  $M$  is a submanifold of  $S$  such that

$$M = \{p(x; \theta(u)) \mid p(x; \theta(u)) \in S, u \in U \subset \mathbf{R}^m\}.$$

**Example 1.1 (normal distributions).** *Let  $S$  be the set of normal distributions,*

$$S = \left\{ p(x; \mu, \sigma) \mid p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right] \right\}.$$

*Here, the sample space  $\mathcal{X}$  is  $\mathbf{R}$ , and the parameter space is the upper half plane  $\Xi = \{(\mu, \sigma) \mid -\infty < \mu < \infty, 0 < \sigma < \infty\}$ .*

*The Fisher metric in  $(\mu, \sigma)$ -coordinate is given by*

$$(g_{ij}^F) = \frac{1}{\sigma^2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

*Hence  $S$  is a space of constant negative curvature  $-1/2$ .*

*Let us change parameters as follows:*

$$\theta^1 = \frac{\mu}{\sigma^2}, \quad \theta^2 = -\frac{1}{2\sigma^2}.$$

*Set*

$$\begin{aligned} Z(x) &= 0, \quad F_1(x) = x, \quad F_2(x) = x^2, \\ \psi(\theta) &= \frac{\mu^2}{2\sigma^2} + \log(\sqrt{2\pi}\sigma) = -\frac{(\theta^1)^2}{4\theta^2} + \frac{1}{2} \log \left( -\frac{\pi}{\theta^2} \right), \end{aligned}$$

*then we obtain*

$$\begin{aligned} p(x; \mu, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right] \\ &= \exp \left[ \frac{\mu}{\sigma^2} x - \frac{1}{2\sigma^2} x^2 - \frac{\mu^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma) \right] \\ &= \exp [x\theta^1 + x^2\theta^2 - \psi(\theta)]. \end{aligned}$$

*This implies that the set of normal distributions is an exponential family.*

For an exponential family, the Fisher metric and the cubic form in  $\{\theta^i\}$ -coordinate are given by

$$g_{ij}^F(\theta) = \partial_i \partial_j \psi(\theta), \tag{1}$$

$$C_{ijk}^F(\theta) = \partial_i \partial_j \partial_k \psi(\theta). \tag{2}$$

The *expectation parameters*  $\{\eta_i\}$  are given by  $\eta_i = E[F_i(x)]$ , and  $\{\eta_i\}$  is a  $\nabla^{(-1)}$ -affine coordinate system.



## 1.2. Statistical manifolds

Let  $(M, h)$  be a semi-Riemannian manifold, and let  $\nabla$  be a torsion-free affine connection on  $M$ . We say that the triplet  $(M, \nabla, h)$  is a *statistical manifold* if  $\nabla h$  is a totally symmetric  $(0, 3)$ -tensor field. Obviously, a statistical model has many statistical manifold structures.

For a statistical manifold  $(M, \nabla, h)$ , we define the *dual connection*  $\nabla^*$  with respect to  $h$  by

$$Xh(Y, Z) = h(\nabla_X Y, Z) + h(Y, \nabla_X^* Z).$$

The connection  $\nabla^*$  is torsion-free and  $\nabla^* h$  is also symmetric. Hence the triplet  $(M, \nabla^*, h)$  is a statistical manifold. We call  $(M, \nabla^*, h)$  the *dual statistical manifold* of  $(M, \nabla, h)$ .

**Proposition 1.2.** *Let  $(M, h)$  be a semi-Riemannian manifold and let  $C$  be a totally symmetric  $(0, 3)$ -tensor field. Denote by  $\nabla^{(0)}$  the Levi-Civita connection  $\nabla^{(0)}$  with respect to  $h$ . We define an affine connection  $\nabla^{(\alpha)}$  by*

$$h(\nabla_X^{(\alpha)} Y, Z) := h(\nabla_X^{(0)} Y, Z) - \frac{\alpha}{2} C(X, Y, Z).$$

*Then, the connections  $\nabla^{(\alpha)}$  and  $\nabla^{(-\alpha)}$  are torsion-free affine connections mutually dual with respect to  $h$ , and the covariant derivative  $\nabla^{(\alpha)} h$  is totally symmetric. Hence  $(M, \nabla^{(\alpha)}, h)$  and  $(M, \nabla^{(-\alpha)}, h)$  are statistical manifolds.*

The connection  $\nabla$  is flat if and only if  $\nabla^*$  is flat. In this case, we say that  $(M, h, \nabla, \nabla^*)$  is a *dually flat space*. Since the connection  $\nabla$  is flat, there exists an affine coordinate system  $\{\theta^i\}$  on  $M$ . In addition, there exists a  $\nabla^*$ -affine coordinate system  $\{\eta_i\}$  such that

$$h\left(\frac{\partial}{\partial \theta^i}, \frac{\partial}{\partial \eta_j}\right) = \delta_i^j.$$

We say that  $\{\eta_i\}$  is the *dual coordinate system* of  $\{\theta^i\}$  with respect to  $h$ .

**Proposition 1.3.** *Let  $(M, h, \nabla, \nabla^*)$  be a dually flat space. Suppose that  $\{\theta^i\}$  is a  $\nabla$ -affine coordinate system, and  $\{\eta_i\}$  is the dual coordinate system of  $\{\theta^i\}$ . Then there exist functions  $\psi$  and  $\phi$  on  $M$  such that*

$$\frac{\partial \psi}{\partial \theta^i} = \eta_i, \quad \frac{\partial \phi}{\partial \eta_i} = \theta^i, \quad \psi(p) + \phi(p) - \sum_{i=1}^n \theta^i(p) \eta_i(p) = 0. \quad (3)$$

*In addition, the following formulas hold:*

$$h_{ij} = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}, \quad h^{ij} = \frac{\partial^2 \phi}{\partial \eta_i \partial \eta_j}, \quad (4)$$

where  $(h_{ij})$  is the component matrix of a semi-Riemannian metric  $h$  with respect to  $\{\theta^i\}$ , and  $(h^{ij})$  is the inverse matrix of  $(h_{ij})$ .

The functions  $\psi$  and  $\phi$  are called the  $\theta$ -potential and the  $\eta$ -potential, respectively. The relation (3) is called the *Legendre transformation*. From Equation (4), the semi-Riemannian metric  $h$  is a Hessian metric. Hence we also say that  $(M, \nabla, h)$  is a *Hessian manifold* [15].

**Definition 1.1.** We say that a function  $\rho$  on  $M \times M$  is the (*canonical*) *divergence* on  $(M, h, \nabla, \nabla^*)$  if

$$\rho(p||q) := \psi(p) + \phi(q) - \sum_{i=1}^n \theta^i(p) \eta_i(q), \quad (p, q \in M). \quad (5)$$

We remark that the definition of  $\rho$  is independent of the choice of affine coordinate system on  $M$ .

### 1.3. Generalized conformal relations on statistical manifolds

We give a brief summary of generalized conformal relations on statistical manifolds. Generalized conformal structures on statistical manifolds have been studied in affine differential geometry (see [5, 6, 7, 8]).

**Definition 1.2.** Suppose  $(M, \nabla, h)$  and  $(M, \bar{\nabla}, \bar{h})$  are statistical manifolds. We say that  $(M, \nabla, h)$  and  $(M, \bar{\nabla}, \bar{h})$  are *conformally-projectively equivalent* if there exist two functions  $\kappa$  and  $\lambda$  such that

$$\begin{aligned} \bar{h}(X, Y) &= e^{\kappa+\lambda} h(X, Y), \\ \bar{\nabla}_X Y &= \nabla_X Y - h(X, Y) \text{grad}_h \lambda + d\kappa(Y) X + d\kappa(X) Y, \end{aligned}$$

where  $\text{grad}_h \lambda$  is the gradient vector field of  $\lambda$  with respect to  $h$ .

In particular, for a constant  $\alpha \in \mathbf{R}$ , we say that two statistical manifolds are  $\alpha$ -conformally equivalent if there exists a function  $\lambda$  on  $M$  such that

$$\begin{aligned} \bar{h}(X, Y) &= e^\lambda h(X, Y), \\ \bar{\nabla}_X Y &= \nabla_X Y - \frac{1+\alpha}{2} h(X, Y) \text{grad}_h \lambda + \frac{1-\alpha}{2} \{d\lambda(Y) X + d\lambda(X) Y\}. \end{aligned}$$

A statistical manifold  $(M, \nabla, h)$  is called  $\alpha$ -conformally flat if  $(M, \nabla, h)$  is locally  $\alpha$ -conformally equivalent to some flat statistical manifold.

We remark that the conformal-projective equivalence relation or the  $\alpha$ -conformal equivalence relation are natural generalizations of conformal

equivalence relation for Riemannian manifolds. In fact, suppose that  $(M, g)$  and  $(M, \bar{g})$  are Riemannian manifolds, and  $\nabla^{(0)}$  and  $\bar{\nabla}^{(0)}$  denote their Levi-Civita connections. If  $g$  and  $\bar{g}$  are conformally equivalent, then the following formulas hold.

$$\begin{aligned}\bar{g}(X, Y) &= e^{2\lambda} g(X, Y), \\ \bar{\nabla}_X^{(0)} Y &= \nabla_X^{(0)} Y - h(X, Y) \text{grad}_h \lambda + d\lambda(Y) X + d\lambda(X) Y.\end{aligned}$$

This implies that  $(M, \nabla^{(0)}, g)$  and  $(M, \bar{\nabla}^{(0)}, \bar{g})$  are 0-conformally equivalent.

To describe generalized conformal structures, let us introduce contrast functions. Let  $\rho$  be a function on  $M \times M$ . We define a function on  $M$  by

$$\rho[X_1 \cdots X_i | Y_1 \cdots Y_j](p) = (X_1)_p \cdots (X_i)_p (Y_1)_q \cdots (Y_j)_q \rho(p|q)|_{p=q},$$

where  $X_1, \cdots, X_i, Y_1, \cdots, Y_j$  are arbitrary vector fields on  $M$ . We call  $\rho$  a *contrast function* on  $M$  if

$$\begin{aligned}\rho(p|p) &= 0 \quad (p \in M), \\ \rho[X] &= \rho[X] = 0, \\ h(X, Y) &:= -\rho[X|Y] \quad \text{is a semi-Riemannian metric on } M.\end{aligned}$$

We remark that the canonical divergence on a dually flat space is a typical example of contrast function.

For a given contrast function  $\rho$  on  $M$ , we can define a torsion-free affine connection by the following formula:

$$h(\nabla_X Y, Z) := -\rho[XY|Z].$$

The triplet  $(M, \nabla, h)$  is a statistical manifold. We say that  $(M, \nabla, h)$  is *induced* from the contrast function  $\rho$ . If we exchange the arguments as  $\rho^*(p|q) := \rho(q|p)$ , then  $\rho^*$  is also a contrast function and induces the dual statistical manifold  $(M, \nabla^*, h)$ . For geometry of contrast functions, the following results are known ([7, 8]).

**Proposition 1.4.** *Let  $\rho$  and  $\bar{\rho}$  be contrast functions on  $M$ , and let  $\lambda$  be a function on  $M$ . Suppose that  $(M, \nabla, h)$  and  $(M, \bar{\nabla}, \bar{h})$  are statistical manifolds induced from  $\rho$  and  $\bar{\rho}$ , respectively.*

- (1) *If  $\bar{\rho}(p|q) = e^{\lambda(p)} \rho(p|q)$ , then two statistical manifolds  $(M, \nabla, h)$  and  $(M, \bar{\nabla}, \bar{h})$  are  $(-1)$ -conformally equivalent.*
- (2) *If  $\bar{\rho}(p|q) = e^{\lambda(q)} \rho(p|q)$ , then two statistical manifolds  $(M, \nabla, h)$  and  $(M, \bar{\nabla}, \bar{h})$  are 1-conformally equivalent.*

## 2. Geometry for $q$ -exponential families

In this section, we discuss geometry of  $q$ -exponential families. A  $q$ -exponential family is a generalization of the standard exponential family. We will consider conformal relations between the standard information geometry and the  $q$ -Fisher geometry.

### 2.1. The $q$ -escort probability and the $q$ -expectation

To begin with, we review the notion of the escort probability and the  $q$ -expectation. Suppose that  $p(x)$  is a probability distribution on  $\mathcal{X}$ . For a fixed number  $q$ , we define the  $q$ -escort distribution  $P_q(x)$  of  $p(x)$  by

$$P_q(x) := \frac{1}{\Omega_q(p)} p(x)^q, \quad \Omega_q(p) := \int_{\mathcal{X}} p(x)^q dx.$$

Let  $f(x)$  be a random variable on  $\mathcal{X}$ . The  $q$ -expectation of  $f(x)$  is the expectation with respect to the  $q$ -escort distribution, that is,

$$E_{q,p}[f(x)] := \int_{\mathcal{X}} f(x) P_q(x) dx = \frac{1}{\Omega_q(p)} \int_{\mathcal{X}} f(x) p(x)^q dx.$$

If the sample space  $\mathcal{X}$  is discrete, the  $q$ -escort distribution or the  $q$ -expectation can be defined by replacing the integral  $\int \cdots dx$  with the sum  $\sum_{x \in \mathcal{X}}$ .

### 2.2. The $q$ -exponential family

Next, we define the  $q$ -exponential and the  $q$ -logarithm. Suppose that  $q$  is a fixed positive number. Then the  $q$ -exponential function is defined by

$$\exp_q x := \begin{cases} (1 + (1 - q)x)^{\frac{1}{1-q}}, & q \neq 1, \quad (1 + (1 - q)x > 0), \\ \exp x, & q = 1, \end{cases} \quad (6)$$

and the  $q$ -logarithm function by

$$\log_q x := \begin{cases} \frac{x^{1-q} - 1}{1-q}, & q \neq 1, \quad (x > 0), \\ \log x, & q = 1. \end{cases}$$

If we consider the limit  $q \rightarrow 1$ , the  $q$ -exponential and the  $q$ -logarithm recover the standard exponential and the standard logarithm, respectively. For simplicity, we assume that the variable  $x$  in (6) satisfy the condition  $1 + (1 - q)x > 0$  if we consider  $q$ -exponential function. Hence  $q$ -exponential and  $q$ -logarithm function are always mutually inverse functions.

**Definition 2.1.** A statistical model  $S_q = \{p(x, \theta) \mid \theta \in \Theta \subset \mathbf{R}^n\}$  is called a  $q$ -exponential family if

$$S_q := \left\{ p(x, \theta) \mid p(x; \theta) = \exp_q \left[ \sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) \right] \right\},$$

where  $F_1(x), \dots, F_n(x)$  are random variables on the sample space  $\mathcal{X}$ , and  $\psi(\theta)$  is a function on the parameter space  $\Theta$ .

The information geometric structure of the  $q$ -exponential family is closely related to the  $(1 - 2q)$ - and the  $(2q - 1)$ -connections. Hence we fix the relations of two parameters  $q$  and  $\alpha$  as  $1 - 2q = \alpha$ .

**Example 2.1 ( $q$ -normal distributions).** A  $q$ -normal distribution is the probability distribution defined by the following formula:

$$p(x; \mu, \sigma) = \frac{1}{Z_{q,\sigma}} \left[ 1 - \frac{1-q}{3-q} \frac{(x - \mu)^2}{\sigma^2} \right]_+^{\frac{1}{1-q}},$$

where  $[*]_+ = \max\{0, *\}$ ,  $\{\mu, \sigma\}$  are parameters  $-\infty < \mu < \infty, 0 < \sigma < \infty$ , and  $Z_{q,\sigma}$  is the normalization defined by

$$Z_{q,\sigma} = \begin{cases} \frac{\sqrt{3-q}}{\sqrt{1-q}} \text{Beta} \left( \frac{2-q}{1-q}, \frac{1}{2} \right) \sigma, & (-\infty < q < 1), \\ \frac{\sqrt{3-q}}{\sqrt{q-1}} \text{Beta} \left( \frac{3-q}{2(q-1)}, \frac{1}{2} \right) \sigma, & (1 \leq q < 3). \end{cases}$$

Set

$$\begin{aligned} \theta^1 &= \frac{2}{3-q} Z_{q,\sigma}^{q-1} \cdot \frac{\mu}{\sigma^2}, \\ \theta^2 &= -\frac{1}{3-q} Z_{q,\sigma}^{q-1} \cdot \frac{1}{\sigma^2}, \\ \psi(\theta) &= -\frac{(\theta^1)^2}{4\theta^2} - \frac{Z_{q,\sigma}^{q-1} - 1}{1-q}, \end{aligned}$$

then

$$\begin{aligned} \log_q p_q(x) &= \frac{1}{1-q} (p^{1-q} - 1) \\ &= \frac{1}{1-q} \left\{ \frac{1}{Z_{q,\sigma}^{1-q}} \left( 1 - \frac{1-q}{3-q} \frac{(x - \mu)^2}{\sigma^2} \right) - 1 \right\} \\ &= \frac{2\mu Z_{q,\sigma}^{q-1}}{(3-q)\sigma^2} x - \frac{Z_{q,\sigma}^{q-1}}{(3-q)\sigma^2} x^2 - \frac{Z_{q,\sigma}^{q-1}}{3-q} \cdot \frac{\mu^2}{\sigma^2} + \frac{Z_{q,\sigma}^{q-1} - 1}{1-q} \\ &= \theta^1 x + \theta^2 x^2 - \psi(\theta). \end{aligned}$$

This implies that the set of  $q$ -normal distributions is a  $q$ -exponential family.

We remark that  $q$ -normal distributions include several important probability distributions. If  $q = 1$ , then the  $q$ -normal distribution is the normal distribution, of course. If  $q = 2$ , then the distribution is the Cauchy distribution. If  $q = 1 + 1/(n+1)$ , then the distribution is Student's  $t$ -distribution. We also remark that mathematical properties of  $q$ -normal distributions have been obtained by several authors. See [16, 17], for example.

**Example 2.2 (discrete distributions).** *Suppose that the sample space  $\mathcal{X}$  is a finite discrete set. Then the set of all probability distributions on  $\mathcal{X}$  is given by*

$$S_n = \left\{ p(x, \eta) \left| \eta_i > 0, \sum_{i=1}^{n+1} \eta_i = 1, p(x; \eta) = \sum_{i=1}^{n+1} \eta_i \delta_i(x) \right. \right\},$$

where  $\delta_i(x)$  equals one if  $x = i$  and zero otherwise. Set

$$\theta^i = \frac{1}{1-q} \{ (\eta_i)^{1-q} - (\eta_{n+1})^{1-q} \},$$

$$\psi(\theta) = -\log_q \eta_{n+1},$$

then we obtain

$$\begin{aligned} \log_q p_q(x) &= \frac{1}{1-q} \{ p^{1-q}(x) - 1 \} \\ &= \frac{1}{1-q} \left\{ \sum_{i=1}^{n+1} (\eta_i)^{1-q} \delta_i(x) - 1 \right\} \\ &= \frac{1}{1-q} \left\{ \sum_{i=1}^n ((\eta_i)^{1-q} - (\eta_{n+1})^{1-q}) \delta_i(x) + (\eta_{n+1})^{1-q} - 1 \right\} \\ &= \sum_{i=1}^n \theta^i \delta_i(x) - \psi(\theta). \end{aligned}$$

This implies that the set of discrete distributions is a  $q$ -exponential family. We note that this also holds in the case  $q = 1$ , that is, the set of discrete distribution is an exponential family.

### 2.3. Geometry for $q$ -exponential families

For a  $q$ -exponential family  $S_q = \{p(x; \theta)\}$ , we assume that the potential function  $\psi$  is strictly convex. We define the  $q$ -Fisher metric and the  $q$ -cubic form in the same manner as exponential families (1) and (2):

$$g_{ij}^q(\theta) = \partial_i \partial_j \psi(\theta),$$

$$C_{ijk}^q(\theta) = \partial_i \partial_j \partial_k \psi(\theta).$$

Since  $g^q$  is a Hessian metric on  $\{S_q\}$ , we can define a flat affine connection  $\nabla^{q(e)} = \nabla^{q(1)}$  by

$$g^q(\nabla_X^{q(e)} Y, Z) = g^q(\nabla_X^{q(0)} Y, Z) - \frac{1}{2} C^q(X, Y, Z),$$

where  $\nabla^{q(0)}$  is the Levi-Civita connection with respect to the  $q$ -Fisher metric  $g^q$ . In this case, the parameters  $\{\theta^i\}$  is a  $\nabla^{q(e)}$ -affine coordinate system. We denote by  $\nabla^{q(m)}$  the dual connection of  $\nabla^{q(e)}$  with respect to  $g^q$ . We call  $\nabla^{q(e)}$  the  $q$ -exponential connection and  $\nabla^{q(m)}$  the  $q$ -mixture connection.

Since  $\nabla^{q(e)}$  is flat, then  $\nabla^{q(m)}$  is also flat. Hence we immediately obtain the following proposition.

**Proposition 2.1.** *Let  $S_q$  be a  $q$ -exponential family. Then the tetrad  $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$  is a dually flat space.*

Let  $S_q$  be a  $q$ -exponential family. From a direct calculation, we have

$$\partial_i p(x; \theta) = p(x; \theta)^q (F_i(x) - \partial_i \psi(\theta)),$$

where  $\partial_i = \partial / \partial \theta^i$ . Since  $\int_{\mathcal{X}} \partial_i p(x, \theta) dx = \partial_i \int_{\mathcal{X}} p(x, \theta) dx = 0$ , we obtain

$$\partial_i \psi(\theta) = \frac{1}{\Omega_q(p)} \int_{\mathcal{X}} F_i(x) p(x; \theta)^q dx = \int_{\mathcal{X}} F_i(x) P_q(x) dx.$$

This implies that the  $q$ -mixture parameters are given by the  $q$ -expectation of the random variables  $\{F_i\}$ . Hence we conclude

**Proposition 2.2.** *Let  $S_q$  be a  $q$ -exponential family. Then the  $q$ -mixture parameters  $\{\eta_i\}$  are given by the  $q$ -expectation of the random variables  $F_i(x)$ , that is,*

$$\eta_i = \frac{\partial}{\partial \theta^i} \psi(\theta) = \int_{\mathcal{X}} F_i(x) P_q(x; \theta) dx.$$

Next, we consider relations between the standard Fisher structure and the  $q$ -Fisher structure from the viewpoint of contrast functions.

For a  $q$ -exponential distribution  $S_q$ , we denote by  $\rho_q$  the canonical divergence (5).

**Proposition 2.3.** *Let  $S_q$  be a  $q$ -exponential family. Then the canonical divergence  $\rho_q$  on  $S_q$  is given by*

$$\rho_q(p(\theta') || p(\theta)) = E_{q, p(\theta)} [\log_q p(\theta) - \log_q p(\theta')].$$

**Proof.** Since  $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$  is a dually flat space, the  $q$ -Fisher metric has a potential function  $\psi$ . We denote  $\phi$  by the dual potential function of  $\psi$ . For probability distributions  $p(\theta)$  and  $p(\theta')$  in  $S_q$ , using the Legendre duality (3), we obtain

$$\begin{aligned}
& E_{q,p(\theta)}[\log_q p(\theta) - \log_q p(\theta')] \\
&= \int_{\mathcal{X}} \left( \sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) - \sum_{i=1}^n (\theta')^i F_i(x) + \psi(\theta') \right) P_q(x; \theta) dx \\
&= \sum_{i=1}^n \theta^i \eta_i - \psi(\theta) - \sum_{i=1}^n (\theta')^i \eta_i + \psi(\theta') \\
&= \psi(\theta') + \phi(\theta) - \sum_{i=1}^n (\theta')^i \eta_i \\
&= \rho_q(p(\theta') || p(\theta)). \quad \square
\end{aligned}$$

We remark that the canonical divergence  $\rho_q(p(\theta) || p(\theta'))$  induces the statistical manifold  $(S_q, \nabla^{q(e)}, g^q)$  and the dual divergence  $\rho_q^*(p(\theta) || p(\theta')) := \rho_q(p(\theta') || p(\theta))$  induces  $(S_q, \nabla^{q(m)}, g^q)$ . The  $q$ -exponential family also has another divergence, called the *divergence of Csiszár type*  $\rho_q^C$ , which is defined by

$$\rho_q^C(p(\theta) || p(\theta')) := \frac{1}{1-q} \left\{ 1 - \int_{\mathcal{X}} p(\theta)^q p(\theta')^{1-q} dx \right\}.$$

This is essentially equivalent to the  $q$  times of the  $(1-2q)$ -divergence in information geometry. The divergence  $(1/q)\rho_q^C$  induces the statistical manifold  $(S_q, \nabla^{(1-2q)}, g^F)$ .

**Proposition 2.4.** *Suppose that  $\rho_q$  and  $\rho_q^C$  are the canonical divergence and the divergence of Csiszár type on a  $q$ -exponential family, respectively. Denote by  $\Omega_q(p(\theta))$  the normalization for the  $q$ -escort distribution of  $p(\theta)$ . Then  $\rho_q$  and  $\rho_q^C$  satisfy*

$$\rho_q(p(\theta') || p(\theta)) = \frac{1}{\Omega_q(p(\theta))} \rho_q^C(p(\theta) || p(\theta')).$$

**Proof.** From Proposition 2.3 we obtain

$$\begin{aligned}
\rho_q(p(\theta') || p(\theta)) &= E_{q,p(\theta)}[\log_q p(\theta) - \log_q p(\theta')] \\
&= \int_{\mathcal{X}} \left( \frac{p(\theta)^{1-q} - 1}{1-q} - \frac{p(\theta')^{1-q} - 1}{1-q} \right) \frac{p(\theta)^q}{\Omega_q(p(\theta))} dx
\end{aligned}$$



$$\begin{aligned}
 &= \frac{1 - \int_{\mathcal{X}} p(\theta)^q p(\theta')^{1-q} dx}{(1-q)\Omega_q(p(\theta))} \\
 &= \frac{1}{\Omega_q(p(\theta))} \rho_q^C(p(\theta) || p(\theta')). \quad \square
 \end{aligned}$$

**Theorem 2.1.** *For a  $q$ -exponential family  $\{S_q\}$ , statistical manifolds  $(S_q, \nabla^{q(e)}, g^q)$  and  $(S_q, \nabla^{(2q-1)}, g^F)$  are 1-conformally equivalent.*

**Proof.** Recall that  $\rho_q(p(\theta) || p(\theta'))$  induces  $(S_q, \nabla^{q(e)}, g^q)$ . From duality of contrast function,  $(1/q)\rho_q^{C*}(p(\theta) || p(\theta')) = (1/q)\rho_q^C(p(\theta') || p(\theta))$  induces  $(S_q, \nabla^{(2q-1)}, g^F)$ . From Proposition 2.4, we have

$$\rho_q(p(\theta) || p(\theta')) = \frac{1}{\Omega_q(p(\theta'))} \rho_q^C(p(\theta') || p(\theta)) = \frac{1}{\Omega_q(p(\theta'))} \rho_q^{C*}(p(\theta) || p(\theta')).$$

This implies that two statistical manifolds are 1-conformally equivalent from Proposition 1.4.  $\square$

We remark that this theorem was already obtained in the case that the sample space  $\mathcal{X}$  is discrete ([13, 14]). For the dual statistical manifolds, we obtain the following corollary immediately.

**Corollary 2.1.** *For a  $q$ -exponential family  $\{S_q\}$ , two statistical manifolds  $(S_q, \nabla^{q(m)}, g^q)$  and  $(S_q, \nabla^{(1-2q)}, g^F)$  are  $(-1)$ -conformally equivalent.*

Since  $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$  is dually flat, we also obtain the following corollary.

**Corollary 2.2.** *For a  $q$ -exponential family  $\{S_q\}$ , the statistical manifold  $(S_q, \nabla^{(2q-1)}, g^F)$  is 1-conformally flat, and  $(S_q, \nabla^{(1-2q)}, g^F)$  is  $(-1)$ -conformally flat.*

For generalization of exponential families, several results have been obtained in more generalized frameworks (see [4, 10, 11, 12]). If we consider relations between the standard Fisher geometry and dually flat structures for them as in our paper, some suitable assumptions may be required.

### 3. An application to statistical inferences

In this section, we discuss an application of geometry of  $q$ -exponential families to statistical inferences along the author's explanatory report [9].

### 3.1. Generalization of independence

At first, let us recall the independence of random variables. Suppose that  $X$  and  $Y$  are random variables which belong to probability density functions  $p_1(x)$  and  $p_2(y)$ , respectively. We say that  $X$  and  $Y$  are *independent* if the joint probability density function  $p(x, y)$  is defined by the product of the marginal probability density functions, that is,

$$p(x, y) = p_1(x)p_2(y).$$

We assume that  $p_1(x)$  and  $p_2(y)$  are positive everywhere on the sample space. Then the above equation can be written as follows:

$$p(x, y) = p_1(x)p_2(y) = \exp [\log p_1(x) + \log p_2(y)].$$

This implies that the notion of independence depends on the duality of the exponential function and the logarithm function, or the law of exponents. Hence we can generalize the notion of independence from the viewpoint of  $q$ -exponential functions.

For a fixed positive number  $q$ , we assume that  $x > 0, y > 0$  and  $x^{1-q} + y^{1-q} - 1 > 0$ . The  $q$ -product [2] of  $x$  and  $y$  is defined by

$$x \otimes_q y := [x^{1-q} + y^{1-q} - 1]^{\frac{1}{1-q}}.$$

The following properties follow from the definition of  $q$ -product.

$$\begin{aligned} \exp_q x \otimes_q \exp_q y &= \exp_q (x + y), \\ \log_q (x \otimes_q y) &= \log_q x + \log_q y. \end{aligned}$$

Let us define the notion of  $q$ -independence. We say that  $X$  and  $Y$  are  $q$ -independent with  $m$ -normalization (mixture normalization) if the joint probability density function  $p_q(x, y)$  is defined by the  $q$ -product of the marginal probability density functions, that is,

$$p_q(x, y) = \frac{p_1(x) \otimes_q p_2(y)}{Z_{p_1, p_2}},$$

where  $Z_{p_1, p_2}$  is the normalization defined by

$$Z_{p_1, p_2} = \int \int_{\mathcal{X} \mathcal{Y}} p_1(x) \otimes_q p_2(y) dx dy.$$

Since the  $q$ -product of probability density functions  $p_1(x) \otimes_q p_2(y)$  is not a probability density in general, a suitable normalization is required [4].

### 3.2. Geometry for $q$ -likelihood estimators

Let  $S = \{p(x; \xi) | \xi \in \Xi\}$  be a statistical model, and let  $\{x_1, \dots, x_N\}$  be  $N$ -independent observations generated from a probability density function  $p(x; \xi) \in S$ . We define the  $q$ -likelihood function [16]  $L_q(\xi)$  by

$$L_q(\xi) = p(x_1; \xi) \otimes_q p(x_2; \xi) \otimes_q \cdots \otimes_q p(x_N; \xi).$$

In the case  $q \rightarrow 1$ , the  $q$ -likelihood function  $L_q$  is the standard likelihood function on  $\Xi$ . Though  $L_q$  may not be a probability density on  $\Xi$ , we regard  $L_q$  as a generalization of the likelihood function.

Since  $q$ -logarithm functions are strictly increasing, it is equivalent to consider the  $q$ -logarithm  $q$ -likelihood function [3]

$$\log_q L_q(\xi) = \sum_{i=1}^N \log_q p(x_i; \xi).$$

We say that  $\hat{\xi}$  is the *maximum  $q$ -likelihood estimator* if

$$\hat{\xi} = \arg \max_{\xi \in \Xi} L_q(\xi) \quad \left( = \arg \max_{\xi \in \Xi} \log_q L_q(\xi) \right).$$

Now let us consider  $q$ -likelihood estimator for  $q$ -exponential families. Let  $S_q$  be a  $q$ -exponential family and let  $M$  be a curved  $q$ -exponential family in  $S$ . Suppose that  $\{x_1, \dots, x_N\}$  are  $N$ -independent observations generated from  $p(x; u) = p(x; \theta(u)) \in M$ .

Then the  $q$ -likelihood function is calculated as

$$\begin{aligned} \log_q L_q(u) &= \sum_{j=1}^N \log_q p(x_j; u) = \sum_{j=1}^N \left\{ \sum_{i=1}^n \theta^i(u) F_i(x_j) - \psi(\theta(u)) \right\} \\ &= \sum_{i=1}^n \theta^i(u) \sum_{j=1}^N F_i(x_j) - N\psi(\theta(u)). \end{aligned}$$

The  $q$ -logarithm  $q$ -likelihood equation is

$$\partial_i \log_q L_q(u) = \sum_{j=1}^N F_i(x_j) - N\partial_i \psi(\theta(u)) = 0.$$

Thus, the  $q$ -likelihood estimator for  $S$  is given by

$$\hat{\eta}_i = \frac{1}{N} \sum_{j=1}^N F_i(x_j).$$

On the other hand, the canonical divergence can be calculated as

$$\begin{aligned}\rho_q^*(p(\hat{\eta})||p(\theta(u))) &= \rho_q(p(\theta(u))||p(\hat{\eta})) \\ &= \psi(\theta(u)) + \phi(\hat{\eta}) - \sum_{i=1}^n \theta^i(u) \hat{\eta}_i \\ &= \phi(\hat{\eta}) - \frac{1}{N} \log_q L_q(u).\end{aligned}$$

Hence the  $q$ -likelihood is maximum if and only if the canonical divergence is minimum. In the same arguments as the standard exponential families, we can say that the  $q$ -likelihood estimator is the orthogonal projection from  $\hat{\eta}$  to the model distribution  $M$  with respect to  $\nabla^{q(m)}$ -geodesic. Hence the  $q$ -likelihood estimator is a quite natural generalization of the likelihood estimator from the viewpoint of differential geometry.

We remark that the  $q$ -likelihood can be generalized by  $U$ -geometry. The notion of independence is related to geometric structures on the sample space [4].

## Acknowledgment

The authors wish to express their sincere gratitude to the referee for his carefully reading and for his apropos comments of the paper.

The first named author is partially supported by The Toyota Physical and Chemical Research Institute and by Grant-in-Aid for Encouragement of Young Scientists (B) No. 19740033, Japan Society for the Promotion of Science.

## References

1. S. Amari and H. Nagaoka, *Methods of information geometry*, Amer. Math. Soc., Providence, Oxford University Press, Oxford, 2000.
2. E.P. Borges, *A possible deformed algebra and calculus inspired in nonextensive thermostatics*, Phys. A, **340**(2004), 95–101.
3. D. Ferrari and Y. Yang, *Maximum Lq-likelihood estimation*, Ann. Statist. **38**(2010), 753–783.
4. Y. Fujimoto and N. Murata, *A Generalization of Independence in Naive Bayes Model*, Lecture Notes in Computer Science, **6283**(2010), 153–161.
5. T. Kurose, *Conformal-projective geometry of statistical manifolds*, Interdiscip. Inform. Sci., **8**(2002), 89–100.
6. H. Matsuzoe, *On realization of conformally-projectively flat statistical manifolds and the divergences*, Hokkaido Math. J., **27**(1998), 409–421.
7. H. Matsuzoe, *Geometry of contrast functions and conformal geometry*, Hiroshima Math. J., **29**(1999), 175–191.

8. H. Matsuzoe, *Computational Geometry from the Viewpoint of Affine Differential Geometry*, Lecture Notes in Computer Science **5416**(2009), 103–123.
9. H. Matsuzoe, *Geometry for statistical inferences in complex systems*, Toyota Research Report, **63**(2011), 177–180.
10. J. Naudts, *Estimators, escort probabilities, and  $\phi$ -exponential families in statistical physics*, JIPAM. J. Inequal. Pure Appl. Math., **5**(2004), Article 102 (electronic).
11. J. Naudts, *Generalised exponential families and associated entropy functions*, Entropy, **10**(2008), 131–149.
12. J. Naudts, *Generalised Thermostatistics*, Springer, 2011.
13. A. Ohara, H. Matsuzoe and S. Amari, *A dually flat structure on the space of escort distributions*, J. Phys.: Conf. Ser. **201**(2010), No. 012012 (electronic).
14. A. Ohara, H. Matsuzoe and S. Amari, *Dually flat structure with escort probability and its application to alpha-Voronoi diagrams*, preprint, arXiv:1010.4965 [stat-mech].
15. H. Shima, *The Geometry of Hessian Structures*, World Scientific, 2007.
16. H. Suyari and M. Tsukada, *Law of Error in Tsallis Statistics*, IEEE Trans. Inform. Theory, **51**(2005), 753–757.
17. M. Tanaka, *Meaning of an escort distribution and  $\tau$ -transformation*, J. Phys.: Conf. Ser. **201**(2010), No 012007 (electronic).
18. C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, Springer, New York, 2009.

Received January 31, 2011

Revised April 16, 2011

# A Hessian domain constructed with a foliation by 1-conformally flat statistical manifolds

by  
Keiko UOHASHI

1

**Abstract.** A Hessian domain is a flat statistical manifold, and its level surfaces are 1-conformally flat statistical submanifolds. In this paper we show conditions that 1-conformally flat statistical leaves of a foliation can be realized as level surfaces of their common Hessian domain conversely.

## 1. Introduction

Let  $\varphi$  be a function on a domain  $\Omega$  in a real affine space  $\mathbf{A}^{n+1}$ . Denoting by  $D$  the canonical flat affine connection on  $\mathbf{A}^{n+1}$ , we set  $g = Dd\varphi$  and suppose that  $g$  is non-degenerate. Then a Hessian domain  $(\Omega, D, g)$  is a flat statistical manifold [8].

Kurose defined  $\alpha$ -conformal equivalence and  $\alpha$ -conformal flatness of statistical manifolds [4]. In [9] we proved that  $n$ -dimensional level surfaces of  $\varphi$  are 1-conformally flat statistical submanifolds of  $(\Omega, D, g)$ . In addition we show properties of foliations on Hessian domains with respect to statistical submanifolds in [10]. Hao and Shima studied level surfaces on Hessian domains deeply in [2] [7]. However they studied foliations and statistical submanifolds for given Hessian domains. We see few results of the realization of statistical manifolds on Hessian domains. In [9] we show that a 1-conformally flat statistical manifold can be locally realized as a submanifold of a flat statistical manifold, constructing a level surface of a Hessian domain. However we proved realization of only "a" 1-conformally flat statistical manifold. In this paper we give conditions for realization of 1-conformally flat statistical manifolds as level surfaces of their common Hessian domain.

---

<sup>1</sup> 2010 *Mathematics Subject Classification.* 53A15

*Key words.* Hessian domain, level surface, foliation, statistical manifold, conformally flat.

In section 2 we recall properties of Hessian domains, statistical manifolds and affine differential geometry. In section 3 we prove a theorem on realization of 1-conformally flat statistical leaves. In section 4 we show necessity of the conditions described in the theorem.

## 2. Hessian domains and Statistical manifolds

Let  $D$  and  $\{x^1, \dots, x^{n+1}\}$  be the canonical flat affine connection and the canonical affine coordinate system on  $\mathbf{A}^{n+1}$ , i.e.,  $Ddx^i = 0$ . If the Hessian  $Dd\varphi = \sum_{i,j} (\partial^2 \varphi / \partial x^i \partial x^j) dx^i dx^j$  is non-degenerate for a function  $\varphi$  on a domain  $\Omega$  in  $\mathbf{A}^{n+1}$ , we call  $(\Omega, D, g = Dd\varphi)$  a Hessian domain. For a torsion-free affine connection  $\nabla$  and a pseudo-Riemannian metric  $h$  on a manifold  $N$ , the triple  $(N, \nabla, h)$  is called a statistical manifold if  $\nabla h$  is symmetric. If the curvature tensor  $R$  of  $\nabla$  vanishes,  $(N, \nabla, h)$  is said to be flat. A Hessian domain  $(\Omega, D, g = Dd\varphi)$  is a flat statistical manifold. Conversely, a flat statistical manifold is locally a Hessian domain [1][8].

For a statistical manifold  $(N, \nabla, h)$ , let  $\nabla'$  be an affine connection on  $N$  such that

$$Xh(Y, Z) = h(\nabla_X Y, Z) + h(Y, \nabla'_X Z), \quad \text{for } X, Y \text{ and } Z \in TN,$$

where  $TN$  is the set of all tangent vector fields on  $N$ . The affine connection  $\nabla'$  is torsion free, and  $\nabla' h$  symmetric. Then  $\nabla'$  is called the dual connection of  $\nabla$ , the triple  $(N, \nabla', h)$  the dual statistical manifold of  $(N, \nabla, h)$ , respectively.

Let  $\mathbf{A}_{n+1}^*$  and  $\{x_1^*, \dots, x_{n+1}^*\}$  be the dual affine space of  $\mathbf{A}^{n+1}$  and the dual affine coordinate system of  $\{x^1, \dots, x^{n+1}\}$ , respectively. We define the gradient mapping  $\iota$  from  $\Omega$  to  $\mathbf{A}_{n+1}^*$  by

$$x_i^* \circ \iota = -\frac{\partial \varphi}{\partial x^i},$$

and a flat affine connection  $D'$  on  $\Omega$  by

$$\iota_*(D'_X Y) = D_X^* \iota_*(Y) \quad \text{for } X, Y \in T\Omega,$$

where  $D_X^* \iota_*(Y)$  is covariant derivative along  $\iota$  induced by the canonical flat affine connection  $D^*$  on  $\mathbf{A}_{n+1}^*$ . Then  $(\Omega, D', g)$  is the dual statistical manifold of  $(\Omega, D, g)$ .

For  $\alpha \in \mathbf{R}$ , statistical manifolds  $(N, \nabla, h)$  and  $(N, \bar{\nabla}, \bar{h})$  are said to be  $\alpha$ -conformally equivalent if there exists a function  $\phi$  on  $N$  such that

$$\begin{aligned}\bar{h}(X, Y) &= e^\phi h(X, Y), \\ h(\bar{\nabla}_X Y, Z) &= h(\nabla_X Y, Z) - \frac{1+\alpha}{2} d\phi(Z) h(X, Y) \\ &\quad + \frac{1-\alpha}{2} \{d\phi(X) h(Y, Z) + d\phi(Y) h(X, Z)\}\end{aligned}$$

for  $X, Y$  and  $Z \in TN$ . A statistical manifold  $(N, \nabla, h)$  is called  $\alpha$ -conformally flat if  $(N, \nabla, h)$  is locally  $\alpha$ -conformally equivalent to a flat statistical manifold. Statistical manifolds  $(N, \nabla, h)$  and  $(N, \bar{\nabla}, \bar{h})$  are  $\alpha$ -conformally equivalent if and only if the dual statistical manifolds  $(N, \nabla', h)$  and  $(N, \bar{\nabla}', \bar{h})$  are  $(-\alpha)$ -conformally equivalent. Especially, a statistical manifold  $(N, \nabla, h)$  is 1-conformally flat if and only if the dual statistical manifold  $(N, \nabla', h)$  is  $(-1)$ -conformally flat [4].

Henceforth, we suppose that  $g$  is positive definite.

Let  $\tilde{E}$  be the gradient vector field of  $\varphi$  on  $\Omega$  defined by

$$g(X, \tilde{E}) = d\varphi(X) \quad \text{for } X \in T\Omega,$$

where  $T\Omega$  is the set of all tangent vector fields on  $\Omega$ . We set

$$E = -d\varphi(\tilde{E})^{-1} \tilde{E} \quad \text{on } \Omega_o = \{p \in \Omega \mid d\varphi_p \neq 0\}.$$

For  $p \in \Omega_o$ ,  $E_p$  is perpendicular to  $T_p M$  with respect to  $g$ , where  $M \subset \Omega_o$  is a level surface of  $\varphi$  containing  $p$  and  $T_p M$  is the set of all tangent vectors at  $p$  on  $M$ .

Let  $x$  be a canonical immersion of an  $n$ -dimensional level surface  $M$  into  $\Omega$ . For  $D$  and an affine immersion  $(x, E)$ , we can define the induced affine connection  $D^E$ , the affine fundamental form  $g^E$  on  $M$  by

$$D_X Y = D_X^E Y + g^E(X, Y) E \quad \text{for } X, Y \in TM.$$

We denote by  $D^M$  and  $g^M$  the connection and the Riemannian metric on  $M$  induced by  $D$  and  $g$ . Then the triple  $(M, D^M, g^M)$  is the statistical submanifold realized in  $(\Omega, D, g)$ , which coincides with the manifold  $(M, D^E, g^E)$  induced by an affine immersion  $(x, E)$ . This fact leads the next theorem.



**Theorem 2.1.** ([9]) *Let  $M$  be a simply connected  $n$ -dimensional level surface of  $\varphi$  on an  $(n+1)$ -dimensional Hessian domain  $(\Omega, D, g = Dd\varphi)$  with a Riemannian metric  $g$ , and suppose that  $n \geq 2$ . If we consider  $(\Omega, D, g)$  a flat statistical manifold,  $(M, D^M, g^M)$  is a 1-conformally flat statistical submanifold of  $(\Omega, D, g)$ , where we denote by  $D^M$  and  $g^M$  the connection and the Riemannian metric on  $M$  induced by  $D$  and  $g$ .*

Conversely, on realization of a 1-conformally flat statistical manifold we have:

**Theorem 2.2.** ([9]) *An arbitrary 1-conformally flat statistical manifold of  $\dim n \geq 2$  with a Riemannian metric can be locally realized as a submanifold of a flat statistical manifold of  $\dim(n+1)$ .*

### 3. Foliations constructed by 1-conformally flat statistical manifolds

Let  $\mathcal{F}$  be a foliation on a differentiable manifold  $N$  of dimension  $n \geq 2$  and codimension 1, and for a leaf  $M \in \mathcal{F}$  the triple  $(M, \nabla^M, h^M)$  a 1-conformally flat statistical manifold. Suppose that a non-degenerate affine immersion  $(x^M, E^M)$  realizes  $(M, \nabla^M, h^M)$  in  $\mathbf{A}^{n+1}$ , and that a mapping  $x : N \rightarrow \Omega$  defined by  $x(p) = x^M(p)$  for  $p \in M$  is a diffeomorphism, where  $\Omega = \cup_{M \in \mathcal{F}} x^M(M) \subset \mathbf{A}^{n+1}$  is a domain diffeomorphic to  $N$ .

We set  $\iota^M$  is the conormal immersion for  $x^M$ , i.e., denoting by  $\langle a, b \rangle$  a pairing of  $a \in \mathbf{A}_{n+1}^*$  and  $b \in \mathbf{A}^{n+1}$ ,

$$\langle \iota^M(p), Y_p \rangle = 0 \quad \text{for } Y_p \in T_p M, \quad \langle \iota^M(p), E_p^M \rangle = 1$$

for  $p \in M$ , considering  $T_p \mathbf{A}^{n+1}$  with  $\mathbf{A}^{n+1}$ . the immersion  $\iota^M$  satisfies that

$$\langle \iota_*^M(Y), E^M \rangle = 0, \quad \langle \iota_*^M(Y), X \rangle = -h^M(Y, X) \quad \text{for } X, Y \in TM$$

Moreover the conormal immersion  $\iota^M$  is equiaffine, i.e.,

$$D_X E^M = S^{E^M}(X) \in TM \quad \text{for } X \in TM$$

(We call  $S^{E^M}$  the shape operator.) [5] [6] [9]. With notations in this section, we can describe

$$D_X Y = \nabla_X^M Y + h^M(X, Y) E^M \quad \text{for } X, Y \in TM.$$

Then the next theorem holds.

**Theorem 3.1.** *If a foliation  $\mathcal{F}$  satisfies the following conditions, each 1-conformally flat statistical leaf  $(M, \nabla^M, h^M)$  of  $\mathcal{F}$  is realized as a level surface of the common Hessian domain:*

- (i) *a mapping  $E : N \rightarrow \mathbf{A}^{n+1}$  defined by  $E(p) = E^M(p)$  for  $p \in M$  is differentiable;*
- (ii) *a mapping  $\iota : N \rightarrow \Omega^*$  defined by  $\iota(p) = \iota^M(p)$  for  $p \in M$  is a diffeomorphism, where  $\Omega^* = \cup_{M \in \mathcal{F}} \iota^M(M) \subset \mathbf{A}_{n+1}^*$ ;*
- (iii)  *$D_E E = \mu E$  for  $\mu \in \mathbf{R}$ ;*
- (iv)  *$S^{E^M}(X) = -(d\lambda(E) + 1)(X)$  on  $M$ , where  $\lambda$  is a function on  $N$  such that  $e^{\lambda(\hat{p})}\iota(p) = \iota(\hat{p})$ ,  $\hat{p} \in N$  for  $p \in M$ .*

*Proof.* We consider a manifold  $N$  a domain  $\Omega \subset \mathbf{A}^{n+1}$ , and define a metric  $g$  on  $\Omega$  by

$$\begin{aligned} g(Y, X) &= h^M(Y, X), \quad g(E, E) = 1, \\ g(Y, E) &= 0 \quad \text{for } X, Y \in TM \subset T\Omega. \end{aligned}$$

Let us prove that  $(D, g)$  satisfies the Codazzi equation

$$(D_X g)(Y, Z) = (D_Y g)(X, Z) \quad \text{for all } X, Y \text{ and } Z \in T\Omega.$$

In the case of  $X, Y$  and  $Z \in TM$ , we have

$$\begin{aligned} (D_X g)(Y, Z) &= X(g(Y, Z)) - g(D_X Y, Z) - g(Y, D_X Z) \\ &= X(h^M(Y, Z)) - g(\nabla_X^M Y, Z) - g(Y, \nabla_X^M Z) \\ &= (\nabla_X^M h^M)(Y, Z). \end{aligned}$$

Similarly it holds that

$$(D_Y g)(X, Z) = (\nabla_Y^M h^M)(X, Z).$$

Recall the Codazzi equation for an equiaffine immersion  $(x^M, E^M)$ ;

$$(\nabla_X^M h^M)(Y, Z) = (\nabla_Y^M h^M)(X, Z)$$

[6]. Then we have the Codazzi equation for  $(D, g)$ .

In the case of  $X, Y \in TM$  and  $E$  on  $M$ , we have

$$\begin{aligned}(D_X g)(Y, E) &= X(g(Y, E)) - g(h^M(X, Y)E, E) - g(Y, D_X E) \\ &= -h^M(X, Y) - h^M(Y, S^{E^M}(X)).\end{aligned}$$

Similarly it holds that

$$(D_Y g)(X, E) = -h^M(X, Y) - h^M(X, S^{E^M}(Y)).$$

Recall the Ricci equation for an equiaffine immersion  $(x^M, E^M)$ ;

$$h^M(S^{E^M}(X), Y) = h^M(X, S^{E^M}(Y))$$

[6]. Then we have the Codazzi equation

$$(D_X g)(Y, E) = (D_Y g)(X, E).$$

In the case of  $X, Z \in TM$  and  $E$  on  $M$ , similarly we have

$$(D_X g)(E, Z) = -h^M(X, Z) - h^M(S^{E^M}(X), Z).$$

Now recall a property  $\langle \iota_*^M(X), E^M \rangle = 0$ ,  $X \in TM$  and the condition (iii)  $D_E E = \mu E$ . Then we have  $D_E X = 0$  for  $X \in TM$ . In addition, conormal immersions  $\{(\iota^{\hat{M}}, h^{\hat{M}})\}_{\hat{M} \in \mathcal{F}}$  are projectively equivalent and conformally equivalent, and it holds that  $h^{\hat{M}} = e^\lambda h^M$  [6]. Hence for  $p \in M$  the next follows;

$$\begin{aligned}E(g(X, Z))|_p &= E(e^\lambda h^M(X, Z))|_p = (Ee^\lambda)|_p h^M(X, Z) \\ &= (E\lambda)|_p e^{\lambda(p)} h^M(X, Z) = d\lambda(E)|_p h^M(X, Z).\end{aligned}$$

Thus it holds that

$$\begin{aligned}(D_E g)(X, Z) &= E(g(X, Z)) - g(D_E X, Z) - g(X, D_E Z) \\ &= d\lambda(E)|_p h^M(X, Z).\end{aligned}$$

By the condition (iv) we have the Codazzi equation

$$(D_X g)(E, Z) = (D_E g)(X, Z).$$

In the case of  $X \in TM$  and  $E$  on  $M$ , we have

$$(D_X g)(E, E) = X(g(E, E)) - g(g(X, E)E, E) - g(E, g(X, E)E) = 0.$$

Moreover by  $D_E X = 0$  and  $D_E E = \mu E$  it holds that

$$(D_E g)(X, E) = X(g(X, E)) - g(D_E X, E) - g(X, D_E E) = 0.$$

Thus we have the Codazzi equation

$$(D_X g)(E, E) = (D_E g)(X, E).$$

In the case of  $X = Y = E$  and  $Z \in T\Omega$ , clearly we have

$$(D_X g)(Y, Z) = (D_Y g)(X, Z) = (D_E g)(E, Z).$$

Hence  $(D, g)$  satisfies the Codazzi equation. Thus  $g$  is a Hessian metric by Proposition 2.1 on [8]. By the definition of  $g$  we can consider that each leaf  $(M, \nabla^M, h^M)$  of  $\mathcal{F}$  is a level surface of the Hessian domain  $(\Omega, D, g)$ .  $\square$

#### 4. Necessity of the conditions

In this section we show that level surfaces of Hessian domain satisfy the conditions of Theorem 3.1.

Let  $(\Omega, D, g = Dd\varphi)$  be a simply connected  $(n + 1)$ -dimensional Hessian domain, and  $(M, D^M, g^M)$  an  $n$ -dimensional 1-conformally flat statistical submanifold on a level surface  $M$  of  $\varphi$ .

It is clear that a mapping  $E : \Omega \rightarrow \mathbf{A}^{n+1}$  defined by  $E(p) = E^M(p)$  for  $p \in M$  is differentiable, where an immersion  $(x^M, E^M)$  realizes  $(M, D^M, g^M)$  in  $\mathbf{A}^{n+1}$ . It is also clear that the gradient mapping  $\iota : \Omega \rightarrow \Omega^* = \iota(\Omega)$  is a diffeomorphism and coincides with the conormal immersion for  $x^M$  on  $M$ . Thus each level surface  $(M, D^M, g^M)$  satisfies the conditions (i) (ii) of Theorem 3.1.

For proof of the condition (iii), we calculate each  $(D_E g)(E, X)$  and  $(D_X g)(E, E)$  for  $X \in TM$ . By the definitions of the gradient vector field  $\tilde{E}$  for  $g$  and the conormal vector field  $E = -d\varphi(\tilde{E})^{-1}\tilde{E}$ , we have

$$\begin{aligned} (D_E g)(E, X) &= E(g(E, X)) - g(D_E E, X) - g(E, D_E X) \\ &= -g(D_E E, X) - d\varphi(\tilde{E})^{-2}d\varphi(D_{\tilde{E}} X) \\ &= -g(D_E E, X) - d\varphi(\tilde{E})^{-2}(\tilde{E}(d\varphi(X)) - (D_{\tilde{E}} d\varphi)(X)) \\ &= -g(D_E E, X). \end{aligned}$$

In the above we also make use of  $d\varphi(X) = 0$  and  $(D_{\tilde{E}}d\varphi)(X) = g(\tilde{E}, X) = 0$ . Moreover it holds that

$$(D_Xg)(E, E) = X(g(E, E)) - 2g(D_XE, E) = -2g(S^{E^M}(X), E) = 0.$$

From the Codazzi equation for  $(D, g)$ , it follows that

$$(D_Eg)(E, X) = (D_Xg)(E, E) = 0.$$

Thus  $D_EE = \mu E$  for  $\mu \in \mathbf{R}$ . Therefore  $(M, D^M, g^M)$  satisfies the condition (iii) of Theorem 3.1.

**Remark 4.1.** Hao and Shima calculated  $(D_{\tilde{E}}g)(\tilde{E}, X)$  and  $(D_Xg)(\tilde{E}, \tilde{E})$  not for  $(x, E)$  but for  $(x, \tilde{E})$ , and showed that the transversal connection form  $\tau^{\tilde{E}}$  vanishes if and only if  $D_{\tilde{E}}\tilde{E} = \mu\tilde{E}$  [2][8]. We gave the above calculation with their technique.

For proof of the condition (iv), we calculate each  $(D_Xg)(E, Z)$  and  $(D_Eg)(X, Z)$  for  $X, Z \in TM$ . By calculation appeared in proof of Theorem 3.1, we have

$$\begin{aligned} (D_Xg)(E, Z) &= -g(X, Z) - g(S^{E^M}(X), Z) \\ (D_Eg)(X, Z) &= d\lambda(E)|_p h^M(X, Z), \end{aligned}$$

where  $\lambda$  is a function on  $\Omega$  defined similar to  $\lambda$  in Theorem 3.1. From the Codazzi equation for  $(D, g)$ , it follows that

$$(D_Xg)(E, Z) = (D_Eg)(X, Z).$$

Thus  $(M, D^M, g^M)$  satisfies the condition (iv)  $S^{E^M}(X) = -(d\lambda(E) + 1)(X)$ .

We describe necessity of the conditions (i) to (iv) as follows.

**Corollary 4.2.** *Each 1-conformally flat statistical leaf  $(M, \nabla^M, h^M)$  of a foliation  $\mathcal{F}$  is realized as a level surface of the common Hessian domain if and only if the  $\mathcal{F}$  satisfies the conditions (i) to (iv) of Theorem 3.1.*

Last we talk about a projectively flat connection and a dual-projectively flat connection. Kurose and Ivanov proved the next propositions, respectively.

**Proposition 4.3.** ([4]) *A statistical manifold  $(N, \nabla, h)$  is 1-conformally flat if and only if the dual connection  $\nabla'$  is a projectively flat connection with symmetric Ricci tensor.*

**Proposition 4.4.** ([3]) *A statistical manifold  $(N, \nabla, h)$  is 1-conformally flat if and only if  $\nabla$  is a dual-projectively flat connection with symmetric Ricci tensor.*

Thus we can describe Corollary 4.2 as the next.

**Corollary 4.5.** *Let  $\nabla^M$  be a dual-projectively flat connection with symmetric Ricci tensor for all  $M \in \mathcal{F}$ . Then each statistical leaf  $(M, \nabla^M, h^M)$  of a foliation  $\mathcal{F}$  is realized as a level surface of the common Hessian domain if and only if  $\mathcal{F}$  satisfies the conditions (i) to (iv) of Theorem 3.1.*

## References

- [1] S.Amari and H.Nagaoka, Method of information geometry, Amer. Math. Soc., Providence, Oxford University Press, Oxford, (2000).
- [2] J.H.Hao and H.Shima: Level surfaces of non-degenerate functions in  $\mathbf{R}^{n+1}$ , Geometriae Dedicata **50** (1994), 193-204.
- [3] S.Ivanov: On dual-projectively flat affine connections, J. of Geom. **53** (1995), 89-99.
- [4] T.Kurose: On the divergence of 1-conformally flat statistical manifolds, Tôhoku Math.J. **46** (1994), 427-433.

- [5] K.Nomizu and U.Pinkall: On the geometry and affine immersions, Math. Z. **195** (1987), 165-178.
- [6] K. Nomizu and T. Sasaki: *Affine Differential Geometry: Geometry of Affine Immersions*, Cambridge Univ. Press (1994).
- [7] H.Shima: Harmonicity of gradient mapping of level surfaces in a real affine space, Geometriae Dedicata **56** (1995), 177-184.
- [8] H.Shima, The geometry of Hessian Structures, World Sci.,(2007).
- [9] K.Uohashi, A.Ohara and T.Fujii: 1-conformally flat statistical submanifolds, Osaka J. math. **37** (2000), 501-507.
- [10] K.Uohashi, A.Ohara and T.Fujii: Foliations and divergences of flat statistical manifolds, Hiroshima Math. J. **30** (2000), 403-414.

Department of Mechanical Engineering & Intelligent Systems  
 Faculty of Engineering  
 Tohoku Gakuin University  
 Miyagi 985-8537, Japan  
 uohashi@tjcc.tohoku-gakuin.ac.jp

# Harmonic maps relative to $\alpha$ -connections on statistical manifolds

Keiko Uohashi

**Abstract.** In this paper we study harmonic maps relative to  $\alpha$ -connections, and not always relative to Levi-Civita connections, on statistical manifolds. In particular, harmonic maps on  $\alpha$ -conformally equivalent statistical manifolds are discussed, and conditions for harmonicity are given by parameters  $\alpha$  and dimensions  $n$ . As the application we also describe harmonic maps between level surfaces of a Hessian domain with  $\alpha$ -conformally flat connections.

**M.S.C. 2010:** 53A15, 53C43.

**Key words:** harmonic map; statistical manifold; dual connection; conformal transformation; Hessian domain.

## 1 Introduction

Harmonic maps are important to research for geometry, physics, and so on. On the other hand statistical manifolds have been studied in terms of affine geometry, information geometry, statistical mechanics, and so on [1]. In relation to them Shima gave conditions for harmonicity of gradient mappings of level surfaces on a Hessian domain, which is a typical example for a dually flat statistical manifold [7] [8].

Level surfaces on a Hessian domain are known as 1- and  $(-1)$ -conformally flat statistical manifolds for the primal connection and for the dual connection, respectively [10]. Then the gradient mappings are considered harmonic maps relative to the dual connection, i.e., the  $(-1)$ -connection. However Shima investigated harmonic maps on  $n$ -dimensional level surfaces into an  $(n + 1)$ -dimensional dual affine space, and not into the other level surfaces. In addition Nomizu and Sasaki calculated the Laplacian of centro-affine immersions into an affine space, which generate projectively flat statistical manifolds, i.e.,  $(-1)$ -conformally flat statistical manifolds. However they show no harmonic maps between two centro-affine hypersurfaces in [6].

Then we treat harmonic maps relative to  $\alpha$ -connections between  $\alpha$ -conformally equivalent statistical manifolds including the case of  $\alpha = -1, 0$  (The 0-connection means the Levi-Civita connection.). In this paper, existence of non trivial harmonic maps for  $\alpha$ -connections is shown with conditions of  $\alpha$ -parameters and dimensions  $n$ . Finally, we describe harmonic maps between level surfaces of a Hessian domain for  $\alpha$ -conformally flat connections.



## 2 Statistical manifolds and $\alpha$ -conformal equivalence

We recall definitions of terms on statistical manifolds.

For a torsion-free affine connection  $\nabla$  and a pseudo-Riemannian metric  $h$  on a manifold  $N$ , the triple  $(N, \nabla, h)$  is called a statistical manifold if  $\nabla h$  is symmetric. If the curvature tensor  $R$  of  $\nabla$  vanishes,  $(N, \nabla, h)$  is said to be flat.

For a statistical manifold  $(N, \nabla, h)$ , let  $\nabla'$  be an affine connection on  $N$  such that

$$Xh(Y, Z) = h(\nabla_X Y, Z) + h(Y, \nabla'_X Z) \quad \text{for } X, Y \text{ and } Z \in \Gamma(TN),$$

where  $\Gamma(TN)$  is the set of smooth tangent vector fields on  $N$ . The affine connection  $\nabla'$  is torsion free, and  $\nabla' h$  symmetric. Then  $\nabla'$  is called the dual connection of  $\nabla$ , the triple  $(N, \nabla', h)$  the dual statistical manifold of  $(N, \nabla, h)$ , and  $(\nabla, \nabla', h)$  the dualistic structure on  $N$ . The curvature tensor of  $\nabla'$  vanishes if and only if that of  $\nabla$  does, and then  $(\nabla, \nabla', h)$  is called the dually flat structure [1].

For a real number  $\alpha$ , statistical manifolds  $(N, \nabla, h)$  and  $(N, \bar{\nabla}, \bar{h})$  are said to be  $\alpha$ -conformally equivalent if there exists a function  $\phi$  on  $N$  such that

$$(2.1) \quad \bar{h}(X, Y) = e^\phi h(X, Y),$$

$$(2.2) \quad \begin{aligned} h(\bar{\nabla}_X Y, Z) &= h(\nabla_X Y, Z) - \frac{1+\alpha}{2} d\phi(Z)h(X, Y) \\ &\quad + \frac{1-\alpha}{2} \{d\phi(X)h(Y, Z) + d\phi(Y)h(X, Z)\} \end{aligned}$$

for  $X, Y$  and  $Z \in \Gamma(TN)$ . Two statistical manifolds  $(N, \nabla, h)$  and  $(N, \bar{\nabla}, \bar{h})$  are  $\alpha$ -conformally equivalent if and only if the dual statistical manifolds  $(N, \nabla', h)$  and  $(N, \bar{\nabla}', \bar{h})$  are  $(-\alpha)$ -conformally equivalent. A statistical manifold  $(N, \nabla, h)$  is called  $\alpha$ -conformally flat if  $(N, \nabla, h)$  is locally  $\alpha$ -conformally equivalent to a flat statistical manifold [4].

## 3 Harmonic maps for $\alpha$ -conformal equivalence

Let  $(N, \nabla, h)$  and  $(N, \bar{\nabla}, \bar{h})$  be  $\alpha$ -conformally equivalent statistical manifolds of  $\dim n \geq 2$ , and  $\{x^1, \dots, x^n\}$  a local coordinate system on  $N$ . Suppose that  $h$  and  $\bar{h}$  are Riemannian metrics. We set  $h_{ij} = h(\partial/\partial x^i, \partial/\partial x^j)$  and  $[h^{ij}] = [h_{ij}]^{-1}$ . Let  $\pi_{id} : N \rightarrow N$  be the identity map, i.e.,  $\pi_{id}(x) = x$  for  $x \in N$ , and  $\pi_{id*}$  the differential of  $\pi_{id}$ . If cautioning about metrics and connections, we denote by  $\pi_{id} : (N, \nabla, h) \rightarrow (N, \bar{\nabla}, \bar{h})$ .

We define a harmonic map relative to  $(h, \nabla, \bar{\nabla})$  as follows.

**Definition 3.1.** If a tension field  $\tau_{(h, \nabla, \bar{\nabla})}(\pi_{id})$  vanishes, i.e.,  $\tau_{(h, \nabla, \bar{\nabla})}(\pi_{id}) \equiv 0$  on  $N$ , the map  $\pi_{id} : (N, \nabla, h) \rightarrow (N, \bar{\nabla}, \bar{h})$  is said to be a harmonic map relative to  $(h, \nabla, \bar{\nabla})$ , where the tension field is defined by

$$(3.1) \quad \tau_{(h, \nabla, \bar{\nabla})}(\pi_{id}) := \sum_{i,j=1}^n h^{ij} \left\{ \bar{\nabla}_{\frac{\partial}{\partial x^i}} (\pi_{id*}(\frac{\partial}{\partial x^j})) - \pi_{id*}(\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}) \right\} \in \Gamma(\pi_{id}^{-1}TN)$$

$$(3.2) \quad = \sum_{i,j=1}^n h^{ij} (\bar{\nabla}_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} - \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}) \in \Gamma(TN).$$

Then the next theorem holds.

**Theorem 3.1.** *For  $\alpha$ -conformally equivalent statistical manifolds  $(N, \nabla, h)$  and  $(N, \bar{\nabla}, \bar{h})$  of  $\dim n \geq 2$  satisfying (2.1) and (2.2), if  $\alpha = -(n-2)/(n+2)$  or  $\phi$  is a constant function on  $N$ , the identity map  $\pi_{id} : (N, \nabla, h) \rightarrow (N, \bar{\nabla}, \bar{h})$  is a harmonic map relative to  $(h, \nabla, \bar{\nabla})$ .*

*Proof.* By (2.2) and (3.2), for  $k \in \{1, \dots, n\}$  we have

$$\begin{aligned} h(\tau_{(h, \nabla, \bar{\nabla})}(\pi_{id}), \frac{\partial}{\partial x^k}) &= h(\sum_{i,j=1}^n h^{ij} (\bar{\nabla}_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} - \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}), \frac{\partial}{\partial x^k}) \\ &= \sum_{i,j=1}^n h^{ij} \left\{ -\frac{1+\alpha}{2} d\phi(\frac{\partial}{\partial x^k}) h(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}) + \frac{1-\alpha}{2} \{ d\phi(\frac{\partial}{\partial x^i}) h(\frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k}) \right. \\ &\quad \left. + d\phi(\frac{\partial}{\partial x^j}) h(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^k}) \} \right\} \\ &= \sum_{i,j=1}^n h^{ij} \left\{ -\frac{1+\alpha}{2} \frac{\partial \phi}{\partial x^k} h_{ij} + \frac{1-\alpha}{2} (\frac{\partial \phi}{\partial x^i} h_{jk} + \frac{\partial \phi}{\partial x^j} h_{ik}) \right\} \\ &= \left\{ -\frac{1+\alpha}{2} \cdot n \cdot \frac{\partial \phi}{\partial x^k} + \frac{1-\alpha}{2} (\sum_{i=1}^n \frac{\partial \phi}{\partial x^i} \delta_{ik} + \sum_{j=1}^n \frac{\partial \phi}{\partial x^j} \delta_{jk}) \right\} \\ &= (-\frac{1+\alpha}{2} \cdot n + \frac{1-\alpha}{2} \cdot 2) \frac{\partial \phi}{\partial x^k} = -\frac{1}{2} \{ (n+2)\alpha + (n-2) \} \frac{\partial \phi}{\partial x^k}, \end{aligned}$$

where  $\delta_{ij}$  is the Kronecker's delta. Therefore, if  $\tau_{(h, \nabla, \bar{\nabla})}(\pi_{id}) \equiv 0$ , it holds that  $(n+2)\alpha + (n-2) = 0$  or  $\partial \phi / \partial x^k = 0$  for all  $k \in \{1, \dots, n\}$  at each point in  $N$ . Thus we obtain Theorem 3.1.  $\square$

## 4 $\alpha$ -connections on level surfaces of a Hessian domain

In this section we show relations with  $\alpha$ -connections and Hessian domains.

Let  $N$  be a manifold with a dualistic structure  $(\nabla, \nabla', h)$ . For  $\alpha \in \mathbf{R}$ , an affine connection defined by

$$(4.1) \quad \nabla^{(\alpha)} := \frac{1+\alpha}{2} \nabla + \frac{1-\alpha}{2} \nabla'$$

is called an  $\alpha$ -connection of  $(N, \nabla, h)$ . The triple  $(N, \nabla^{(\alpha)}, h)$  is also a statistical manifold, and  $\nabla^{(-\alpha)}$  the dual connection of  $\nabla^{(\alpha)}$ . The 1-connection, the  $(-1)$ -connection and the 0-connection coincide with  $\nabla$ ,  $\nabla'$  and the Levi-Civita connection of  $(N, h)$ , respectively. An  $\alpha$ -connection is not always flat [1].

Let  $D$  and  $\{x^1, \dots, x^{n+1}\}$  be the canonical flat affine connection and the canonical affine coordinate system on  $\mathbf{A}^{n+1}$ , i.e.,  $Ddx^i = 0$ . If the Hessian  $Dd\varphi = \sum_{i,j=1}^{n+1} (\partial^2 \varphi / \partial x^i \partial x^j) dx^i dx^j$  is non-degenerate for a function  $\varphi$  on a domain  $\Omega$  in  $\mathbf{A}^{n+1}$ , we call  $(\Omega, D, g = Dd\varphi)$  a Hessian domain. A Hessian domain is a flat statistical manifold. Conversely, a flat statistical manifold is locally a Hessian domain [1] [8].

Let  $\mathbf{A}_{n+1}^*$  and  $\{x_1^*, \dots, x_{n+1}^*\}$  be the dual affine space of  $\mathbf{A}^{n+1}$  and the dual affine coordinate system of  $\{x^1, \dots, x^{n+1}\}$ , respectively. We define the gradient mapping  $\iota$  from  $\Omega$  to  $\mathbf{A}_{n+1}^*$  by

$$x_i^* \circ \iota = -\frac{\partial \varphi}{\partial x^i},$$

and a flat affine connection  $D'$  on  $\Omega$  by

$$\iota_*(D'_X Y) = D_X^* \iota_*(Y) \quad \text{for } X, Y \in \Gamma(T\Omega),$$

where  $D_X^* \iota_*(Y)$  is covariant derivative along  $\iota$  induced by the canonical flat affine connection  $D^*$  on  $\mathbf{A}_{n+1}^*$ . Then  $(\Omega, D', g)$  is the dual statistical manifold of  $(\Omega, D, g)$  [7] [8].

For a simply connected level surface  $M$  of  $\varphi$  with  $\dim n \geq 2$ , we denote by  $D^M$  and  $g^M$  the connection and the Riemannian metric on  $M$  induced by  $D$  and  $g$ , respectively. Then  $(M, D^M, g^M)$  is a 1-conformally flat statistical submanifold of  $(\Omega, D, g)$  by Theorem 2.1 in [10].

We consider two simply connected level surfaces of  $\dim n \geq 2$   $(M, D, g)$ ,  $(\hat{M}, \hat{D}, \hat{g})$  1-conformally flat statistical submanifolds of  $(\Omega, D, g)$ . For  $p \in M$ , let  $\lambda$  be a function on  $M$  such that  $e^{\lambda(p)} \iota(p) \in \hat{\iota}(\hat{M})$ , where  $\hat{\iota}$  is the restriction of the gradient mapping  $\iota$  to  $\hat{M}$ , and set  $(e^\lambda)(p) = e^{\lambda(p)}$ . Note that the function  $e^\lambda$  means the projection of  $M$  to  $\hat{M}$  with respect to the dual affine coordinate system of  $\Omega$ .

We define a map  $\pi : M \rightarrow \hat{M}$  by

$$\hat{\iota} \circ \pi = e^\lambda \iota,$$

denoting also by  $\iota$  the restriction of the gradient mapping  $\iota$  to  $M$ . We denote by  $\bar{D}'$  an affine connection on  $M$  defined by

$$\pi_*(\bar{D}'_X Y) = \hat{D}'_{\pi_*(X)} \pi_*(Y) \quad \text{for } X, Y \in \Gamma(TM),$$

and by  $\bar{g}$  a Riemannian metric on  $M$  such that

$$\bar{g}(X, Y) = e^\lambda g(X, Y) = \hat{g}(\pi_*(X), \pi_*(Y)).$$

Then the next theorem is known (cf. [4] [5]).

**Theorem 4.1.** ([11]) *For affine connections  $D', \bar{D}'$  on  $M$ , we have*

- (i)  *$D'$  and  $\bar{D}'$  are projectively equivalent.*
- (ii)  *$(M, D', g)$  and  $(M, \bar{D}', \bar{g})$  are  $(-1)$ -conformally equivalent.*

We denote by  $\bar{D}$  an affine connection on  $M$  defined by

$$\pi_*(\bar{D}_X Y) = \hat{D}_{\pi_*(X)} \pi_*(Y) \quad \text{for } X, Y \in \Gamma(TM).$$

From duality of  $\hat{D}$  and  $\hat{D}'$ ,  $\bar{D}$  is the dual connection of  $\bar{D}'$  on  $M$ . Then the next theorem holds (cf. [3] [4]).

**Theorem 4.2.** ([11]) For affine connections  $D, \bar{D}$  on  $M$ , we have

- (i)  $D$  and  $\bar{D}$  are dual-projectively equivalent.
- (ii)  $(M, D, g)$  and  $(M, \bar{D}, \bar{g})$  are 1-conformally equivalent.

For  $\alpha$ -connections  $D^{(\alpha)}$ ,  $\bar{D}^{(\alpha)} = D^{(-\alpha)}$  defined similarly to (4.1), we obtain the next corollary by Theorem 4.1, Theorem 4.2 and by (2.2) with  $\phi = \lambda$  [9].

**Corollary 4.3.** For affine connections  $D^{(\alpha)}$ ,  $\bar{D}^{(\alpha)}$  on  $M$ ,  $(M, D^{(\alpha)}, g)$  and  $(M, \bar{D}^{(\alpha)}, \bar{g})$  are  $\alpha$ -conformally equivalent.

## 5 Harmonic maps relative to $\alpha$ -connections on level surfaces

We denote  $\hat{D}_{\pi_*(X)}^{(\alpha)} \pi_*(Y)$  by  $\hat{D}_X^{(\alpha)} \pi_*(Y)$ , considering it in the induced section  $\Gamma(\pi^{-1}T\hat{M})$ . Let  $\{x^1, \dots, x^n\}$  be a local coordinate system on  $M$ . A harmonic map between level surfaces  $(M, D^{(\alpha)}, g)$  and  $(\hat{M}, \hat{D}^{(\alpha)}, \hat{g})$  is defined as follows.

**Definition 5.1.** If a tension field  $\tau_{(g, D^{(\alpha)}, \hat{D}^{(\alpha)})}(\pi)$  vanishes, i.e.,  $\tau_{(g, D^{(\alpha)}, \hat{D}^{(\alpha)})}(\pi) \equiv 0$  on  $M$ , the map  $\pi : (M, D^{(\alpha)}, g) \rightarrow (\hat{M}, \hat{D}^{(\alpha)}, \hat{g})$  is said to be a harmonic map relative to  $(g, D^{(\alpha)}, \hat{D}^{(\alpha)})$ , where the tension field defined by

$$(5.1) \quad \tau_{(g, D^{(\alpha)}, \hat{D}^{(\alpha)})}(\pi) := \sum_{i,j=1}^n g^{ij} \left\{ \hat{D}_{\frac{\partial}{\partial x^i}}^{(\alpha)} \left( \pi_* \left( \frac{\partial}{\partial x^j} \right) \right) - \pi_* \left( D_{\frac{\partial}{\partial x^i}}^{(\alpha)} \frac{\partial}{\partial x^j} \right) \right\} \in \Gamma(\pi^{-1}T\hat{M}).$$

Now we give conditions for harmonicity of a map  $\pi : M \rightarrow \hat{M}$  relative to  $(g, D^{(\alpha)}, \hat{D}^{(\alpha)})$ .

**Theorem 5.1.** Let  $(M, D^{(\alpha)}, g)$  and  $(\hat{M}, \hat{D}^{(\alpha)}, \hat{g})$  be simply connected  $n$ -dimensional level surfaces of an  $(n+1)$ -dimensional Hessian domain  $(\Omega, D, g)$  with  $n \geq 2$ . If  $\alpha = -(n-2)/(n+2)$  or  $\lambda$  is a constant function on  $M$ , a map  $\pi : (M, D^{(\alpha)}, g) \rightarrow (\hat{M}, \hat{D}^{(\alpha)}, \hat{g})$  is a harmonic map relative to  $(g, D^{(\alpha)}, \hat{D}^{(\alpha)})$ , where

$$\hat{\iota} \circ \pi = e^{\lambda} \iota, \quad (e^{\lambda})(p) = e^{\lambda(p)}, \quad e^{\lambda(p)} \iota(p) \in \hat{\iota}(\hat{M}), \quad p \in M,$$

and  $\iota, \hat{\iota}$  are the restrictions of the gradient mapping on  $\Omega$  to  $M, \hat{M}$ , respectively.

*Proof.* The tension field of the map  $\pi$  relative to  $(g, D^{(\alpha)}, \hat{D}^{(\alpha)})$  is described with  $(M, \bar{D}^{(\alpha)}, \bar{g})$ , which is the pull-back of  $(\hat{M}, \hat{D}^{(\alpha)}, \hat{g})$ , as follows.

$$\begin{aligned} \tau_{(g, D^{(\alpha)}, \hat{D}^{(\alpha)})}(\pi) &= \sum_{i,j=1}^n g^{ij} \left\{ \hat{D}_{\frac{\partial}{\partial x^i}}^{(\alpha)} \left( \pi_* \left( \frac{\partial}{\partial x^j} \right) \right) - \pi_* \left( D_{\frac{\partial}{\partial x^i}}^{(\alpha)} \frac{\partial}{\partial x^j} \right) \right\} \\ &= \sum_{i,j=1}^n g^{ij} \left\{ \pi_* \left( \bar{D}_{\frac{\partial}{\partial x^i}}^{(\alpha)} \frac{\partial}{\partial x^j} \right) - \pi_* \left( D_{\frac{\partial}{\partial x^i}}^{(\alpha)} \frac{\partial}{\partial x^j} \right) \right\} = \pi_* \left( \sum_{i,j=1}^n g^{ij} \left( \bar{D}_{\frac{\partial}{\partial x^i}}^{(\alpha)} \frac{\partial}{\partial x^j} - D_{\frac{\partial}{\partial x^i}}^{(\alpha)} \frac{\partial}{\partial x^j} \right) \right) \end{aligned}$$

Identifying  $T_{\pi(x)}M$  with  $T_xM$ , and considering the definition of  $\pi$ , we have

$$\tau_{(g, D^{(\alpha)}, \hat{D}^{(\alpha)})}(\pi) = e^{\lambda} \sum_{i,j=1}^n g^{ij} \left( \bar{D}_{\frac{\partial}{\partial x^i}}^{(\alpha)} \frac{\partial}{\partial x^j} - D_{\frac{\partial}{\partial x^i}}^{(\alpha)} \frac{\partial}{\partial x^j} \right).$$

By Corollary 4.3,  $(M, D^{(\alpha)}, g)$  and  $(M, \bar{D}^{(\alpha)}, \bar{g})$  are  $\alpha$ -conformally equivalent, so that we have the equation (2.2) with  $\phi = \lambda$ ,  $h = g$ ,  $\nabla = D^{(\alpha)}$  and  $\bar{\nabla} = \bar{D}^{(\alpha)}$  for  $X, Y$  and  $Z \in \Gamma(TM)$ . Then it holds similarly to the proof of Theorem 3.1 that for  $k \in \{1, \dots, n\}$

$$\begin{aligned} g(\tau_{(g, D^{(\alpha)}, \hat{D}^{(\alpha)})}(\pi), \frac{\partial}{\partial x^k}) &= g(e^\lambda \sum_{i,j=1}^n g^{ij} (\bar{D}_{\frac{\partial}{\partial x^i}}^{(\alpha)} \frac{\partial}{\partial x^j} - D_{\frac{\partial}{\partial x^i}}^{(\alpha)} \frac{\partial}{\partial x^j}), \frac{\partial}{\partial x^k}) \\ &= e^\lambda \sum_{i,j=1}^n g^{ij} \left\{ -\frac{1+\alpha}{2} d\lambda \left( \frac{\partial}{\partial x^k} \right) g \left( \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right) + \frac{1-\alpha}{2} \left\{ d\lambda \left( \frac{\partial}{\partial x^i} \right) g \left( \frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k} \right) \right. \right. \\ &\quad \left. \left. + d\lambda \left( \frac{\partial}{\partial x^j} \right) g \left( \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^k} \right) \right\} \right\} \\ &= \left( -\frac{1+\alpha}{2} \cdot n + \frac{1-\alpha}{2} \cdot 2 \right) e^\lambda \frac{\partial \lambda}{\partial x^k} = -\frac{1}{2} \{ (n+2)\alpha + (n-2) \} e^\lambda \frac{\partial \lambda}{\partial x^k}. \end{aligned}$$

Therefore, if  $\tau_{(g, D^{(\alpha)}, \hat{D}^{(\alpha)})}(\pi) \equiv 0$ , it holds that  $(n+2)\alpha + (n-2) = 0$  or  $\partial\lambda/\partial x^k = 0$  for all  $k \in \{1, \dots, n\}$  at each point in  $N$ . Thus we obtain Theorem 5.1.  $\square$

Comparing proofs of Theorem 3.1 and Theorem 5.1, we have the following about two tension fields.

**Corollary 5.2.** *Let  $\pi : (M, D^{(\alpha)}, g) \rightarrow (\hat{M}, \hat{D}^{(\alpha)}, \hat{g})$  be the map defined at Theorem 5.1, and  $\pi_{id} : (M, D^{(\alpha)}, g) \rightarrow (M, \bar{D}^{(\alpha)}, \bar{g})$  the identity map, where  $(M, \bar{D}^{(\alpha)}, \bar{g})$  is the pull-back of  $(\hat{M}, \hat{D}^{(\alpha)}, \hat{g})$  by  $\pi$ . Then it holds that*

$$\tau_{(g, D^{(\alpha)}, \hat{D}^{(\alpha)})}(\pi) = e^\lambda \tau_{(g, D^{(\alpha)}, \bar{D}^{(\alpha)})}(\pi_{id}).$$

**Remark 5.2.** For  $n = 2$ , if and only if  $\alpha = 0$ , there exist harmonic maps  $\pi_{id}$  and  $\pi$  with non constant functions  $\phi$  and  $\lambda$ , respectively.

**Remark 5.3.** For  $n \geq 3$ , it holds that  $-1 < \alpha < 0$  if a map  $\pi_{id}$  or  $\pi$  is a harmonic map with a non constant function  $\phi$  or  $\lambda$ , respectively.

**Remark 5.4.** For  $\alpha \leq -1$  and  $\alpha > 0$ , there exist no harmonic maps  $\pi_{id}$  and  $\pi$  with non constant functions  $\phi$  and  $\lambda$ , respectively.

**Acknowledgement.** The author is partially supported by AOARD Grant No.10-1-4060.

## References

- [1] S. Amari and H. Nagaoka, *Method of information geometry*, Amer. Math. Soc., Providence, Oxford University Press, Oxford 2000.
- [2] J.H. Hao and H. Shima, *Level surfaces of non-degenerate functions in  $\mathbf{R}^{n+1}$* , Geometriae Dedicata 50 (1994), 193-204.
- [3] S. Ivanov, *On dual-projectively flat affine connections*, J. of Geom. 53 (1995), 89-99.

- [4] T. Kurose, *On the divergence of 1-conformally flat statistical manifolds*, Tôhoku Math. J. 46 (1994), 427-433.
- [5] K. Nomizu and U. Pinkall, *On the geometry and affine immersions*, Math. Z. 195 (1987), 165-178.
- [6] K. Nomizu and T. Sasaki, *Affine Differential Geometry: Geometry of Affine Immersions*, Cambridge Univ. Press, Cambridge 1994.
- [7] H. Shima, *Harmonicity of gradient mapping of level surfaces in a real affine space*, Geometriae Dedicata 56 (1995), 177-184.
- [8] H. Shima, *The geometry of Hessian Structures*, World Sci. 2007.
- [9] K. Uohashi, *On  $\alpha$ -conformal equivalence of statistical submanifolds*, J. of Geom. 75 (2002), 179-184.
- [10] K. Uohashi, A. Ohara and T. Fujii, *1-conformally flat statistical submanifolds*, Osaka J. math. 37 (2000), 501-507.
- [11] K. Uohashi, A. Ohara and T. Fujii, *Foliations and divergences of flat statistical manifolds*, Hiroshima Math. J. 30 (2000), 403-414.

*Author's address:*

Keiko Uohashi  
Department of Mechanical Engineering & Intelligent Systems,  
Faculty of Engineering, Tohoku Gakuin University  
Tagajo, Miyagi 985-8537, Japan.  
E-mail: uohashi@tjcc.tohoku-gakuin.ac.jp

## CONFORMAL GEOMETRY OF ESCORT PROBABILITY AND ITS APPLICATIONS

ATSUMI OHARA

*Department of Electrical and Electronics Engineering,  
University of Fukui, Bunkyo 3-9-1, Fukui 910-8507, Japan  
ohara@fuee.u-fukui.ac.jp*

HIROSHI MATSUZOE

*Department of Computer Science and Engineering,  
Graduate School of Engineering, Nagoya Institute of Technology,  
Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan  
matsuzoe@nitech.ac.jp*

SHUN-ICHI AMARI

*Riken Brain Science Institute, Wako-shi Hirosawa 2-1, Saitama 351-0198, Japan  
amari@brain.riken.jp*

Received 11 November 2011

Accepted 5 February 2012

Escort probability is a certain modification of ordinary probability and a conformally transformed structure can be introduced on the space of its distributions. In this contribution applications of escort probabilities and such a structure are focused on. We demonstrate that they naturally appear and play important roles for computationally efficient method to construct  $\alpha$ -Voronoi partitions and analysis of related dynamical systems on the simplex.

*Keywords:* Voronoi partitions; dynamical systems; information geometry.

### 1. Introduction

In the research areas of multifractals and nonextensive statistical mechanics, escort probability<sup>1-3</sup> appears in many aspects and is widely recognized as an important concept. It has been known<sup>4,5</sup> that nonextensive entropies are closely connected with the  $\alpha$ -geometry.<sup>6,7</sup> Further, we have geometrically studied the space of escort distributions and reported<sup>8-10</sup> that the well-established and abundant structure (called the *dually flat structure*) can be introduced by a conformal transformation of the  $\alpha$ -geometry.

The purpose of this contribution is to show that escort probability and the associated conformal structure are also natural and useful to the other applications.

First, we discuss the Voronoi partition with respect to the  $\alpha$ -divergence (or Rényi divergence). The Voronoi partitions on the space of probability distributions with the Kullback–Leibler,<sup>11,12</sup> or Bregman divergences<sup>13</sup> are useful tools for various statistical modeling problems involving pattern classification, clustering, likelihood ratio test and so on. See also the literature<sup>14–16</sup> for related problems. The largest advantage to take account of  $\alpha$ -divergences is their *invariance under transformations by sufficient statistics*,<sup>7,17</sup> which is a significant requirement for those statistical applications. In computational aspect, the conformal flattening of the  $\alpha$ -geometry enables us to invoke the standard algorithm<sup>18,19</sup> using a potential function and an upper envelop of hyperplanes with the escort probabilities as coordinates. As another application, we explore properties of dynamical systems defined by the escort transformation and the gradient with respect to the conformal metric. These flows are fundamental from geometrical viewpoints<sup>20</sup> and found to possess interesting properties.

The paper is organized as follows: Sec. 2 is a short review of properties of information geometric structure induced on the family of escort distributions obtained by the authors.<sup>8</sup> Section 3 describes the first application of escort probability and the conformal geometric structure to  $\alpha$ -Voronoi partitions on the simplex. The properties including computational efficiency of a construction algorithm are discussed. Further, a formula for  $\alpha$ -centroid is touched upon. In Sec. 4, we discuss properties of dynamical systems related with escort transformation and gradient flows in view of the conformal geometry.

In the sequel, we use two equivalent parameters  $q$  and  $\alpha$  following to conventions of several research areas, but their relation is fixed as  $q = (1 + \alpha)/2$ . Additionally, we assume that  $q > 0$ .

## 2. Preliminary Results

In this section, we review and summarize results in Ref. 8.

Let  $\mathcal{S}^n$  denote the  $n$ -dimensional probability simplex, i.e.

$$\mathcal{S}^n := \left\{ \mathbf{p} = (p_i) \left| p_i > 0, \sum_{i=1}^{n+1} p_i = 1 \right. \right\}, \quad (1)$$

and  $p_i, i = 1, \dots, n+1$  denote probabilities of  $n+1$  states. We introduce the  $\alpha$ -geometric structure<sup>6,7</sup> on  $\mathcal{S}^n$ . Let  $\{\partial_i\}, i = 1, \dots, n$  be natural basis tangent vector fields on  $\mathcal{S}^n$  defined by

$$\partial_i := \frac{\partial}{\partial p_i} - \frac{\partial}{\partial p_{n+1}}, \quad i = 1, \dots, n, \quad (2)$$

where  $p_{n+1} = 1 - \sum_{i=1}^n p_i$ . Now we define a Riemannian metric  $g$  on  $\mathcal{S}^n$  called the



Fisher metric:

$$g_{ij}(\mathbf{p}) := g(\partial_i, \partial_j) = \frac{1}{p_i} \delta_{ij} + \frac{1}{p_{n+1}} \\ = \sum_{k=1}^{n+1} p_k (\partial_i \log p_k) (\partial_j \log p_k), \quad i, j = 1, \dots, n. \quad (3)$$

Further, define a torsion-free affine connection  $\nabla^{(\alpha)}$  called the  $\alpha$ -connection, which is represented in its coefficients with a real parameter  $\alpha$  by

$$\Gamma_{ij}^{(\alpha)k}(\mathbf{p}) = \frac{1+\alpha}{2} \left( -\frac{1}{p_k} \delta_{ij}^k + p_k g_{ij} \right), \quad i, j, k = 1, \dots, n, \quad (4)$$

where  $\delta_{ij}^k$  is equal to one if  $i = j = k$  and zero otherwise. Then we have the  $\alpha$ -covariant derivative  $\nabla^{(\alpha)}$ , which gives

$$\nabla_{\partial_i}^{(\alpha)} \partial_j = \sum_{k=1}^n \Gamma_{ij}^{(\alpha)k} \partial_k,$$

when it is applied to the vector fields  $\partial_i$  and  $\partial_j$ . We can define a distance-like function on  $\mathcal{S}^n \times \mathcal{S}^n$  for  $\alpha \neq \pm 1$  by

$$D^{(\alpha)}(\mathbf{p}, \mathbf{r}) = \frac{4}{1-\alpha^2} \left\{ 1 - \sum_{i=1}^{n+1} (p_i)^{(1-\alpha)/2} (r_i)^{(1+\alpha)/2} \right\},$$

which we call the  $\alpha$ -divergence. The Fisher metric  $g$  and the  $\alpha$ -connection  $\nabla^{(\alpha)}$  can be derived from the  $\alpha$ -divergence.<sup>7,21</sup>

Since  $\nabla^{(\alpha)}$  and  $\nabla^{(-\alpha)}$  geometrically play dualistic roles<sup>6,7</sup> with respect to  $g$ , we consider the triple  $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ , which is called the  $\alpha$ -geometric structure on  $\mathcal{S}^n$ . The properties of the Tsallis entropy are studied through the  $\alpha$ -geometry.<sup>4,5</sup>

While the  $\alpha$ -geometric structure for  $\alpha \neq \pm 1$  is not flat, we reported<sup>8</sup> that it can be flattened via a certain conformal transformation<sup>22-25</sup> to a nonstandard dually flat structure<sup>6,7</sup> denoted by  $(h, \nabla, \nabla^*)$ . The theoretical advantage or interesting aspect of such a conformally flattening is that we can obtain the Legendre structure on  $\mathcal{S}^n$  preserving several properties of the  $\alpha$ -geometric structure. We summarize the result in the following proposition by preparing some notation: the *escort probability*<sup>1</sup>  $P_i$  and a function  $Z_q$  are respectively defined for  $q \in \mathbf{R}$  by

$$P_i(\mathbf{p}) := \frac{(p_i)^q}{\sum_{j=1}^{n+1} (p_j)^q}, \quad i = 1, \dots, n+1, \quad Z_q(\mathbf{p}) := \sum_{i=1}^{n+1} \frac{(p_i)^q}{q}. \quad (5)$$

For  $0 < q$  with  $q \neq 1$ , we define two functions by

$$\ln_q(s) := \frac{s^{1-q} - 1}{1-q}, \quad s \geq 0, \quad \exp_q(t) := [1 + (1-q)t]_+^{1/(1-q)}, \quad t \in \mathbf{R},$$

where  $[t]_+ := \max\{0, t\}$ , and the so-called Tsallis entropy<sup>26</sup> by

$$S_q(\mathbf{p}) := \frac{\sum_{i=1}^{n+1} (p_i)^q - 1}{1-q}.$$

Note that  $s = \exp_q(\ln_q(s))$  holds and they respectively recover the usual logarithmic, exponential function and the Boltzmann–Gibbs–Shannon entropy  $-\sum_{i=1}^{n+1} p_i \ln p_i$  when  $q \rightarrow 1$ . For  $q > 0$ ,  $\ln_q(s)$  is concave on  $s > 0$ .

**Proposition 1.** *The dually flat structure  $(h, \nabla, \nabla^*)$  on  $\mathcal{S}^n$  is induced via a conformal transformation from the  $\alpha$ -structure  $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$  on  $\mathcal{S}^n$ . The induced potential functions  $\psi, \psi^*$ , and dually flat affine coordinate systems  $(\theta^1, \dots, \theta^n)$  and  $(\eta_1, \dots, \eta_n)$  are represented as follows:*

$$\begin{aligned}\theta^i(\mathbf{p}) &= \ln_q(p_i) - \ln_q(p_{n+1}), \quad i = 1, \dots, n, \\ \eta_i(\mathbf{p}) &= P_i(\mathbf{p}), \quad i = 1, \dots, n, \\ \psi(\boldsymbol{\theta}(\mathbf{p})) &= -\ln_q(p_{n+1}), \\ \psi^*(\boldsymbol{\eta}(\mathbf{p})) &= \frac{1}{\kappa} (\lambda(\mathbf{p}) - q),\end{aligned}$$

where  $\kappa = (1 - \alpha^2)/4 = q(1 - q)$  is the scalar curvature of the  $\alpha$ -structure,  $\theta^{n+1} \equiv 0$ ,  $\eta_{n+1} := P_{n+1}(\mathbf{p}) = 1 - \sum_{i=1}^n P_i(\mathbf{p})$  and  $\lambda = 1/Z_q$  is a conformal factor, i.e.  $h = \lambda g$ .

Further, the coordinate systems  $(\theta^1, \dots, \theta^n)$  and  $(\eta_1, \dots, \eta_n)$  are  $\nabla$ - and  $\nabla^*$ -affine, respectively.

For the proofs of Proposition 1 and necessary lemmas, see Ref. 27. The result is extended to the  $q$ -exponential family with continuous random variables.<sup>9,10</sup>

Note that by defining what we call the *conformal divergence*  $\rho$ ,

$$\begin{aligned}\rho(\mathbf{p}, \mathbf{r}) &:= \lambda(\mathbf{r}) D^{(\alpha)}(\mathbf{p}, \mathbf{r}) = \sum_{i=1}^{n+1} -P_i(\mathbf{r}) (\ln_q(p_i) - \ln_q(r_i)) \\ &= \psi(\boldsymbol{\theta}(\mathbf{p})) + \psi^*(\boldsymbol{\eta}(\mathbf{r})) - \sum_{i=1}^n \theta^i(\mathbf{p}) \eta_i(\mathbf{r}),\end{aligned}\tag{6}$$

we can confirm the Legendre structure, i.e. relations  $\rho(\mathbf{p}, \mathbf{p}) = 0$ ,  $\forall \mathbf{p} \in \mathcal{S}^n$  and

$$\eta_i = \frac{\partial \psi}{\partial \theta^i}, \quad \theta^i = \frac{\partial \psi^*}{\partial \eta_i}, \quad i = 1, \dots, n.\tag{7}$$

The dual potential  $\psi^*$  can be alternatively represented<sup>8</sup> in  $\mathbf{p}$  by

$$\psi^* = \ln_q \left( \frac{1}{\exp_q(S_q(\mathbf{p}))} \right),$$

which is known as the negative of the *normalized Tsallis entropy*.<sup>28–30</sup> Thus, when  $q \rightarrow 1$ , we have the standard dually flat structure on  $\mathcal{S}^n$  as follows:

$$\psi \rightarrow -\ln p_{n+1}, \quad \psi^* \rightarrow \sum_{i=1}^{n+1} p_i \ln p_i \quad \theta^i \rightarrow \ln(p_i/p_{n+1}), \quad \eta_i \rightarrow p_i, \quad i = 1, \dots, n.$$

Finally, it should be remarked that the both structures  $(h, \nabla, \nabla^*)$  and  $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$  are related in terms of not only the conformality of the metrics  $h = \lambda g$  but also the *projective equivalence*<sup>31</sup> between the connections  $\nabla^*$  and  $\nabla^{(-\alpha)}$ ,<sup>a</sup> which implies that a curve on  $\mathcal{S}^n$  is  $\nabla^*$ -geodesic if and only if it is  $\nabla^{(-\alpha)}$ -geodesic.<sup>b</sup> More generally, a submanifold in  $\mathcal{S}^n$  is  $\nabla^*$ -autoparallel if and only if it is  $\nabla^{(-\alpha)}$ -autoparallel. For  $(h, \nabla, \nabla^*)$ , in particular, a submanifold is  $\nabla$ - (resp.  $\nabla^*$ -) autoparallel when the affine coordinates  $\theta^i$  (resp.  $\eta_i$ ) are affinely parametrized by  $\beta^j, j = 1, \dots, m \leq n$  as  $\theta^i = \sum_{j=1}^m a_j^i \beta^j + c^i$ , for  $i = 1, \dots, n+1$  (similarly for  $\eta_i$ ). For example, the  $q$ -exponential family

$$p_i = \exp_q\{\theta^i - \tilde{\psi}(\boldsymbol{\beta})\}, \quad i = 1, \dots, n+1, \quad (8)$$

where  $\tilde{\psi}$  is a normalizing term defined by  $\tilde{\psi} = \theta^{n+1} + \psi$ , is  $\nabla$ -autoparallel in a proper domain of  $\boldsymbol{\beta}$ . These properties are crucially used in the following sections.

Proposition 1 with (7) implies that

$$P_i = \frac{\partial \psi}{\partial \theta^i}, \quad i = 1, \dots, n \quad (9)$$

for  $p_i = \exp_q(\theta^i - \psi)$ ,  $i = 1, \dots, n$  and  $p_{n+1} = \exp_q(-\psi)$ . This relation can be regarded as a special case of a known one<sup>3,32</sup> for the  $q$ -exponential family (8), using the *escort expectation*,<sup>2</sup>

$$\langle \langle a_j \rangle \rangle_q := \sum_{i=1}^{n+1} P_i a_j^i = \frac{1}{q Z_q} \sum_{i=1}^{n+1} (p_i)^q \frac{\partial}{\partial \beta^j} (\ln_q(p_i) + \tilde{\psi} - c^i) = \frac{\partial \tilde{\psi}}{\partial \beta^j},$$

because (9) is derived when  $a_j^i = \delta_j^i, j = 1, \dots, n$  and  $a_{n+1}^i = c^i = 0$ .

### 3. Applications to Construction of Alpha-Voronoi Partitions and Alpha-Centroids

For given  $m$  points  $\mathbf{p}_1, \dots, \mathbf{p}_m$  on  $\mathcal{S}^n$  we define  $\alpha$ -Voronoi regions on  $\mathcal{S}^n$  using the  $\alpha$ -divergence as follows:

$$\text{Vor}^{(\alpha)}(\mathbf{p}_k) := \bigcap_{l \neq k} \{\mathbf{p} \in \mathcal{S}^n | D^{(\alpha)}(\mathbf{p}_k, \mathbf{p}) < D^{(\alpha)}(\mathbf{p}_l, \mathbf{p})\}, \quad k = 1, \dots, m.$$

An  $\alpha$ -Voronoi partition (diagram) on  $\mathcal{S}^n$  is a collection of the  $\alpha$ -Voronoi regions and their boundaries. Note that  $D^{(\alpha)}$  approaches the Kullback–Leibler (KL) divergence if  $\alpha \rightarrow -1$ , and  $D^{(0)}$  is called the Hellinger distance. If we use the Rényi divergence<sup>33</sup> of order  $\alpha \neq 1$  defined by

$$D_\alpha(\mathbf{p}, \mathbf{r}) := \frac{1}{\alpha - 1} \ln \sum_{i=1}^{n+1} (p_i)^\alpha (r_i)^{1-\alpha}$$

<sup>a</sup>Note that  $\nabla^*$  is projectively equivalent with  $\nabla^{(\alpha)}$  in Ref. 8 because there we adopted a different correspondence of parameters:  $q = (1 - \alpha)/2$ .

<sup>b</sup>Precisely speaking, the term “geodesic” should be replaced by “pre-geodesic”.

instead of the  $\alpha$ -divergence,  $\text{Vor}^{(1-2\alpha)}(\mathbf{p}_k)$  gives the corresponding Voronoi region because of their one-to-one functional relationship.

The standard algorithm using projection of a polyhedron<sup>18,19</sup> commonly works well to construct Voronoi partitions for the Euclidean distance,<sup>19</sup> the KL divergence.<sup>12</sup> The algorithm is generally applicable if a divergence function is of *Bregman type*,<sup>13</sup> which is represented by the remainder of the first order Taylor expansion of a convex potential function in a suitable coordinate system. Geometrically speaking, this implies that i) the divergence is of the form (6) in a dually flat structure and ii) its affine coordinate system is chosen to realize the corresponding Voronoi partitions. In this coordinate system with one extra complementary coordinate the polyhedron is expressed as the upper envelop of  $m$  hyperplanes tangent to the potential function.

A problem for the case of the  $\alpha$ -Voronoi partition is that the  $\alpha$ -divergence on  $\mathcal{S}^n$  cannot be represented as a remainder of any convex potentials. The following theorem, however, claims that the problem is resolved by Proposition 1, i.e. conformally transforming the  $\alpha$ -geometry to the dually flat structure  $(h, \nabla, \nabla^*)$  and using the conformal divergence  $\rho$  and escort probabilities as a coordinate system.

Here, we denote the space of escort distributions by  $\mathcal{E}^n$  and represent the point on  $\mathcal{E}^n$  by  $\mathbf{P} = (P_1, \dots, P_n)$  because  $P_{n+1} = 1 - \sum_{i=1}^n P_i$ .

**Theorem 1.**

- (i) The bisector of  $\mathbf{p}_k$  and  $\mathbf{p}_l$  defined by  $\{\mathbf{p} | D^{(\alpha)}(\mathbf{p}_k, \mathbf{p}) = D^{(\alpha)}(\mathbf{p}_l, \mathbf{p})\}$  is a simultaneously  $\nabla^{(-\alpha)}$ - and  $\nabla^*$ -autoparallel hypersurface on  $\mathcal{S}^n$ .
- (ii) Let  $\mathcal{H}_k, k = 1, \dots, m$  be the hyperplane in  $\mathcal{E}^n \times \mathbf{R}$  which is respectively tangent at  $(\mathbf{P}_k, \psi^*(\mathbf{P}_k))$  to the hypersurface  $\{(\mathbf{P}, y) | y = \psi^*(\mathbf{P})\}$ , where  $\mathbf{P}_k = \mathbf{P}(\mathbf{p}_k)$ . The  $\alpha$ -Voronoi diagram can be constructed on  $\mathcal{E}^n$  as the projection of the upper envelope of  $\mathcal{H}_k$ 's along the  $y$ -axis.

**Proof.** (i) Consider the  $\nabla^{(\alpha)}$ -geodesic  $\gamma^{(\alpha)}$  connecting  $\mathbf{p}_k$  and  $\mathbf{p}_l$ , and let  $\bar{\mathbf{p}}$  be the midpoint on  $\gamma^{(\alpha)}$  satisfying  $D^{(\alpha)}(\mathbf{p}_k, \bar{\mathbf{p}}) = D^{(\alpha)}(\mathbf{p}_l, \bar{\mathbf{p}})$ . Denote by  $\mathcal{B}$  the  $\nabla^{(-\alpha)}$ -autoparallel hypersurface that is orthogonal to  $\gamma^{(\alpha)}$  and passes  $\bar{\mathbf{p}}$ . Then, for all  $\mathbf{r} \in \mathcal{B}$ , the modified Pythagorean theorem<sup>4,23</sup> implies the following equality:

$$\begin{aligned} D^{(\alpha)}(\mathbf{p}_k, \mathbf{r}) &= D^{(\alpha)}(\mathbf{p}_k, \bar{\mathbf{p}}) + D^{(\alpha)}(\bar{\mathbf{p}}, \mathbf{r}) - \kappa D^{(\alpha)}(\mathbf{p}_k, \bar{\mathbf{p}}) D^{(\alpha)}(\bar{\mathbf{p}}, \mathbf{r}) \\ &= D^{(\alpha)}(\mathbf{p}_l, \bar{\mathbf{p}}) + D^{(\alpha)}(\bar{\mathbf{p}}, \mathbf{r}) - \kappa D^{(\alpha)}(\mathbf{p}_l, \bar{\mathbf{p}}) D^{(\alpha)}(\bar{\mathbf{p}}, \mathbf{r}) = D^{(\alpha)}(\mathbf{p}_l, \mathbf{r}). \end{aligned}$$

Hence,  $\mathcal{B}$  is a bisector of  $\mathbf{p}_k$  and  $\mathbf{p}_l$ . The projective equivalence ensures that  $\mathcal{B}$  is also  $\nabla^*$ -autoparallel.

(ii) Recall the conformal relation (6) between  $D^{(\alpha)}$  and  $\rho$ , then we see that  $\text{Vor}^{(\alpha)}(\mathbf{p}_k) = \text{Vor}^{(\text{conf})}(\mathbf{p}_k)$  holds on  $\mathcal{S}^n$ , where

$$\text{Vor}^{(\text{conf})}(\mathbf{p}_k) := \bigcap_{l \neq k} \{\mathbf{p} \in \mathcal{S}^n | \rho(\mathbf{p}_k, \mathbf{p}) < \rho(\mathbf{p}_l, \mathbf{p})\}.$$

Proposition 1 and the Legendre relations (6) and (7) imply that  $\rho(\mathbf{p}_k, \mathbf{p})$  is represented with the coordinates  $(P_i)$  by

$$\rho(\mathbf{p}_k, \mathbf{p}) = \psi^*(\mathbf{P}) - \left( \psi^*(\mathbf{P}_k) + \sum_{i=1}^n \frac{\partial \psi^*}{\partial P_i}(\mathbf{P}_k)(P_i(\mathbf{p}) - P_i(\mathbf{p}_k)) \right),$$

where  $\mathbf{P} = \mathbf{P}(\mathbf{p})$ . Note that a point  $(\mathbf{P}, y_k(\mathbf{P}))$  in  $\mathcal{H}_k$  is expressed by

$$y_k(\mathbf{P}) := \psi^*(\mathbf{P}_k) + \sum_{i=1}^n \frac{\partial \psi^*}{\partial P_i}(\mathbf{P}_k)(P_i(\mathbf{p}) - P_i(\mathbf{p}_k)).$$

Hence, we have  $\rho(\mathbf{p}_k, \mathbf{p}) = \psi^*(\mathbf{P}) - y_k(\mathbf{P})$ . We see, for example, that the bisector on  $\mathcal{E}^n$  for  $\mathbf{p}_k$  and  $\mathbf{p}_l$  is represented as a projection of  $\mathcal{H}_k \cap \mathcal{H}_l$ . Thus, the statement follows.  $\square$

Figures 1 and 2 taken from Ref. 27 show examples of  $\alpha$ -Voronoi partitions for four common probability distributions on  $\mathcal{S}^2$ : (0.2, 0.7, 0.1), (0.3, 0.3, 0.4), (0.4, 0.4, 0.2), (0.6, 0.1, 0.3) with  $\alpha = -0.6$  and 2. While the left ones are represented with usual probabilities on  $\mathcal{S}^2$  (the axis  $p_3$  is omitted), right ones are the corresponding partitions represented with escort probabilities on  $\mathcal{E}^2$ . In right ones of the both figures, the bisectors are straight line segments on  $\mathcal{E}^2$  because they are simultaneously  $\nabla^{(-\alpha)}$ - and  $\nabla^*$ -geodesics as is proved in (i) of Theorem 1.

**Remark 1.** Voronoi partitions for broader class of divergences that are not necessarily associated with any convex potentials are theoretically studied<sup>34</sup> from more general affine differential geometric points of views.

On the other hand, the  $\alpha$ -divergence can be expressed as a Bregman divergence if the domain is extended from  $\mathcal{S}^n$  to the positive orthant  $\mathbf{R}_+^{n+1}$ .<sup>5-7</sup> Hence, the  $\alpha$ -geometry on  $\mathbf{R}_+^{n+1}$  is dually flat. Using this property,  $\alpha$ -Voronoi partitions on  $\mathbf{R}_+^{n+1}$  is discussed by Nielsen and Nock.<sup>35</sup>

However, while both of the above mentioned methods require constructions of the polyhedrons in the space of dimension  $d = n + 2$ , the new one proposed in this paper does in the space of dimension  $d = n + 1$ . Since it is known<sup>36</sup> that the optimal computational time of polyhedrons depends on the dimension  $d$  by  $O(m \log m + m^{\lfloor d/2 \rfloor})$ , the new one is better when  $n$  is even and  $m$  is large.

The next proposition is a simple and relevant application of escort probabilities. Define the  $\alpha$ -centroid  $\mathbf{c}^{(\alpha)}$  for given  $m$  points  $\mathbf{p}_1, \dots, \mathbf{p}_m$  on  $\mathcal{S}^n$  by the minimizer of the following problem:

$$\min_{\mathbf{p} \in \mathcal{S}^n} \sum_{k=1}^m D^{(\alpha)}(\mathbf{p}, \mathbf{p}_k).$$

**Proposition 2.** The  $\alpha$ -centroid  $\mathbf{c}^{(\alpha)}$  for given  $m$  points  $\mathbf{p}_1, \dots, \mathbf{p}_m$  on  $\mathcal{S}^n$  is represented in escort probabilities by the weighted average of conformal factors

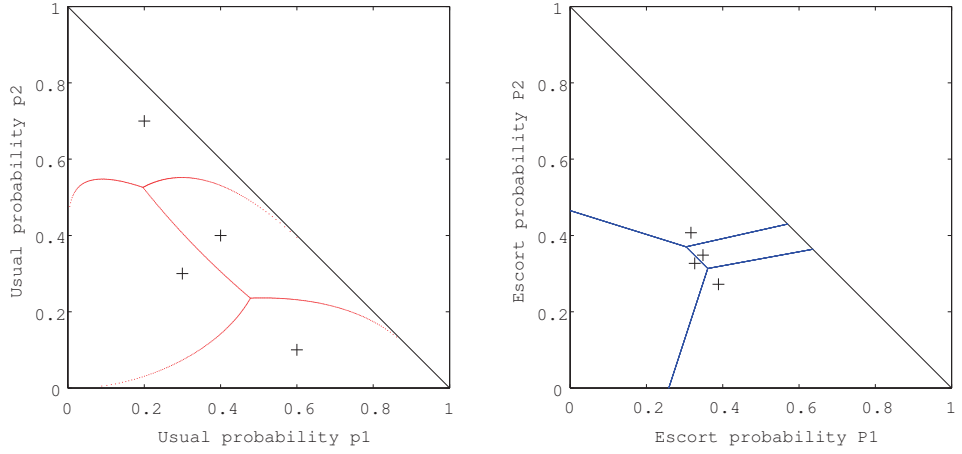


Fig. 1. An example of  $\alpha$ -Voronoi partition on  $\mathcal{S}^2$  (left) for  $\alpha = -0.6$  (or  $q = 0.2$ ) and the corresponding one on  $\mathcal{E}^2$  (right).

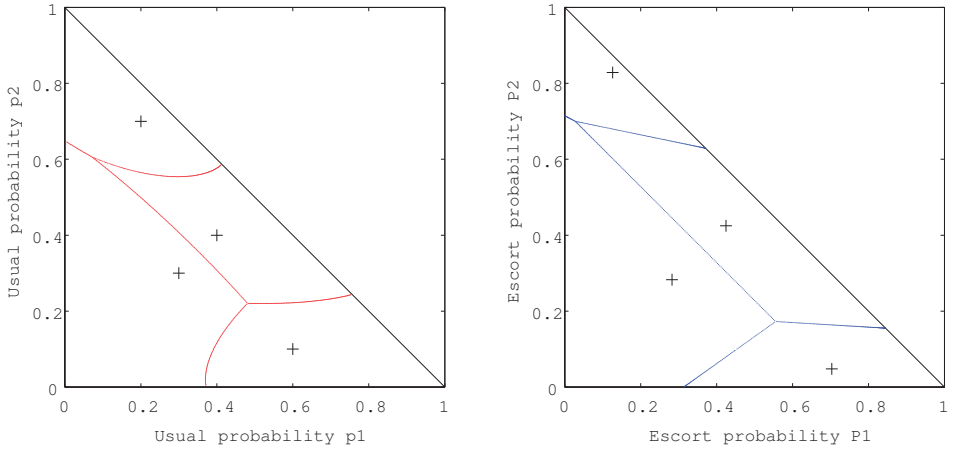


Fig. 2. An example of  $\alpha$ -Voronoi partition on  $\mathcal{S}^2$  (left) for  $\alpha = 2$  (or  $q = 1.5$ ) and the corresponding one on  $\mathcal{E}^2$  (right).

$\lambda(\mathbf{p}_k) = 1/Z_q(\mathbf{p}_k)$ , i.e.

$$P_i(c^{(\alpha)}) = \frac{1}{\sum_{k=1}^m Z_q(\mathbf{p}_k)} \sum_{k=1}^m Z_q(\mathbf{p}_k) P_i(\mathbf{p}_k), \quad i = 1, \dots, n+1.$$

**Proof.** Let  $\theta^i = \theta^i(\mathbf{p})$ . Using (6), we have

$$\sum_{k=1}^m D^{(\alpha)}(\mathbf{p}, \mathbf{p}_k) = \sum_{k=1}^m Z_q(\mathbf{p}_k) \rho(\mathbf{p}, \mathbf{p}_k) = \sum_{k=1}^m Z_q(\mathbf{p}_k) \left\{ \psi(\boldsymbol{\theta}) + \psi^*(\boldsymbol{\eta}(\mathbf{p}_k)) - \sum_{i=1}^n \theta^i \eta_i(\mathbf{p}_k) \right\}.$$

Then the optimality condition is

$$\frac{\partial}{\partial \theta^i} \sum_{k=1}^m D^{(\alpha)}(\mathbf{p}, \mathbf{p}_k) = \sum_{k=1}^m Z_q(\mathbf{p}_k)(\eta_i - \eta_i(\mathbf{p}_k)) = 0, \quad i = 1, \dots, n,$$

where  $\eta_i = \eta_i(\mathbf{p})$ . Thus, the statements for  $i = 1, \dots, n$  follow from Proposition 1. For  $i = n + 1$ , it holds since the sum of the weights is equal to one.  $\square$

#### 4. Related Dynamical Systems on the Simplex

In this section, we study properties of several dynamical systems naturally associated with the escort transformation, the conformal flattening and the resultant geometric structure.

##### 4.1. Conformal replicator equation

Recall the *replicator system* on the simplex  $\mathcal{S}^n$  for given functions  $f_i(\mathbf{p})$  defined by

$$\dot{p}_i = p_i(f_i(\mathbf{p}) - \bar{f}(\mathbf{p})), \quad i = 1, \dots, n+1, \quad \bar{f}(\mathbf{p}) := \sum_{i=1}^{n+1} p_i f_i(\mathbf{p}), \quad (10)$$

which is extensively studied in evolutionary game theory. It is known<sup>37</sup> that

- (i) the solution of (10) is the gradient flow of a function  $V(\mathbf{p})$  satisfying

$$f_i = \frac{\partial V}{\partial p_i}, \quad i = 1, \dots, n+1,$$

with respect to the *Shahshahani metric*,<sup>38</sup>

- (ii) the KL divergence is a local Lyapunov function for an equilibrium called the *evolutionary stable state (ESS)*.

The Shahshahani metric is defined on the positive orthant  $\mathbf{R}_+^{n+1}$  by

$$\tilde{g}_{ij} = \frac{\sum_{k=1}^{n+1} p_k}{p_i} \delta_{ij}, \quad i, j = 1, \dots, n+1.$$

Note that a vector  $X = \sum_{i=1}^n X^i \partial_i$  tangent to  $\mathcal{S}^n$  is represented by a tangent vector  $\tilde{X}$  on  $\mathbf{R}_+^{n+1}$  by  $\tilde{X} = \sum_{k=1}^{n+1} \tilde{X}^k \partial/\partial p_k$ , where  $\tilde{X}^i = X^i$ ,  $i = 1, \dots, n$  and  $\tilde{X}^{n+1} = -\sum_{i=1}^n X^i$ . Then we see that the Shahshahani metric induces the Fisher metric  $g$  in (3) on  $\mathcal{S}^n$  because  $\sum_{i,j}^n g_{ij} X^i X^j = \sum_{k,l}^{n+1} \tilde{g}_{kl} \tilde{X}^k \tilde{X}^l$  holds. Further, the KL divergence is a canonical divergence<sup>7</sup> of  $(g, \nabla^{(1)}, \nabla^{(-1)})$ . Thus, the replicator dynamics (10) are closely related with the standard dually flat structure  $(g, \nabla^{(1)}, \nabla^{(-1)})$ , which associates with exponential and mixture families of probability distributions.<sup>39</sup>

In this subsection, motivated by the above two features (i) and (ii), we define a modified replicator system compatible to the dually flat structure  $(h, \nabla, \nabla^*)$  and discuss their properties. See Harper<sup>40</sup> for another modification of the replicator system.

Consider a metric on  $\mathbf{R}_+^{n+1}$  defined by  $\tilde{h} := \lambda \tilde{g}$  and the following modified replicator system:

$$\dot{p}_i = Z_q(\mathbf{p}) p_i (f_i(\mathbf{p}) - \bar{f}(\mathbf{p})), \quad i = 1, \dots, n+1. \quad (11)$$

It is easy to see the above right-hand sides define the vector that is tangent to  $\mathcal{S}^n$  and the gradient of a function  $V$  with respect to  $\tilde{h}$ , since  $\sum_{i=1}^{n+1} \dot{p}_i = 0$  and

$$\tilde{h}(\tilde{X}, \dot{\mathbf{p}}) = \sum_{i,j=1}^{n+1} \tilde{h}_{ij} \tilde{X}^i \dot{p}_j = \sum_{i=1}^{n+1} f_i \tilde{X}^i - \bar{f} \sum_{i=1}^{n+1} \tilde{X}^i = \sum_{i=1}^{n+1} \frac{\partial V}{\partial p_i} \tilde{X}^i,$$

respectively, hold for any tangent vector  $\tilde{X}$  on  $\mathcal{S}^n$ . Thus, comparing (10) and (11), we can conclude as follows:

**Proposition 3.** *The gradient flow of a function  $V$  on  $\mathcal{S}^n$  with respect to the conformal metric  $h$  is given by (11). Its trajectories coincide with those of (10) while velocities of time-evolutions are different by the factor  $Z_q(\mathbf{p})$ .*

We investigate properties of (11) in the case that  $V(\mathbf{p}) = -\rho(\mathbf{r}, \mathbf{p})$  for a fixed distribution  $\mathbf{r}$ . Applying the result for gradient flows of divergences on dually flat spaces,<sup>20</sup> we see that the flow is explicitly given in the  $\nabla$ -affine coordinates by

$$\theta^i(\mathbf{p}(t)) = \exp(-t) \{ \theta^i(\mathbf{p}(0)) - \theta^i(\mathbf{r}) \} + \theta^i(\mathbf{r}), \quad i = 1, \dots, n, \quad (12)$$

i.e. it converges to  $\mathbf{r}$  along the  $\nabla$ -geodesic (pregeodesic) curve.

On the other hand, consider the optimization problem maximizing  $V(\mathbf{p}) = -\rho(\mathbf{r}, \mathbf{p})$  with  $m$  constraints of the escort expectations:

$$\begin{aligned} \langle \langle A_j \rangle \rangle_q &= \sum_{i=1}^{n+1} P_i(\mathbf{p}) A_j^i \\ &= \sum_{i=1}^n \eta_i(\mathbf{p}) A_j^i + \left( 1 - \sum_{i=1}^n \eta_i(\mathbf{p}) \right) A_j^{n+1} = \bar{A}_j, \quad j = 1, \dots, m, \end{aligned} \quad (13)$$

where  $A_j^i$  and  $\bar{A}_j$  are prescribed values. Since the constraints (13) form a  $\nabla^*$ -autoparallel submanifold in  $\mathcal{S}^n$ , the problem has the *unique* maximizer owing to the Pythagorean theorem<sup>6,7</sup> in a dually flat space. Defining the Lagrangian

$$L(\mathbf{p}) := \rho(\mathbf{r}, \mathbf{p}) + \sum_{j=1}^m \beta^j (\bar{A}_j - \langle \langle A_j \rangle \rangle_q),$$

we have the following optimality condition from (6) and (7):

$$\begin{aligned} \frac{\partial L}{\partial \eta_i} &= \theta^i - \theta_{\mathbf{r}}^i - \sum_{j=1}^m \beta^j (A_j^i - A_j^{n+1}) \\ &= \ln_q p_i + \psi(\boldsymbol{\theta}) - \theta_{\mathbf{r}}^i - \sum_{j=1}^m \beta^j (A_j^i - A_j^{n+1}) = 0, \quad i = 1, \dots, n, \end{aligned}$$



where  $\theta^i$  and  $\eta_i$  are, respectively, the  $\nabla$ - and the  $\nabla^*$ -affine coordinates of  $\mathbf{p}$  introduced in Theorem 1, and  $\theta_{\mathbf{r}}^i := \theta^i(\mathbf{r})$ . Hence,  $\theta^i$  is affine with respect to  $\beta^j$  and the maximizer  $\mathbf{p}$  is in the  $q$ -exponential family represented in (8). These facts imply that the set of maximizers forms a  $\nabla$ -autoparallel submanifold parametrized by  $\beta^j$ , which are determined by the prescribed values  $\bar{A}_j$ .

Combining this consideration with (12), we see that the following holds:

**Corollary 1.** *Let  $\mathbf{r}$  be any distribution, and suppose that  $\mathbf{p}_0$  and  $\mathbf{p}_\infty$  are in the  $q$ -exponential family (8) parametrized by  $\beta^j$  as  $\theta^i = \sum_{j=1}^m (A_j^i - A_j^{n+1})\beta^j + \theta_{\mathbf{r}}^i$ ,  $i = 1, \dots, n$  and  $\theta^{n+1} \equiv 0$ . The gradient flow (11) with  $V(\mathbf{p}) = -\rho(\mathbf{p}_\infty, \mathbf{p})$  starting from  $\mathbf{p}_0$  converges to  $\mathbf{p}_\infty$  staying on the  $q$ -exponential family.*

In the above,  $\mathbf{p}_0$  and  $\mathbf{p}_\infty$  are respectively interpreted as maximizers of  $-\rho(\mathbf{r}, \mathbf{p})$  under the constraints (13) with different values of  $\bar{A}_j$ 's. The corollary claims that the  $q$ -exponential family is an invariant manifold for the transition of distribution from  $\mathbf{p}_0$  to  $\mathbf{p}_\infty$  caused by the change of  $\bar{A}_j$ 's, if the transition dynamics are governed by the gradient flow.

#### 4.2. Flows of escort transformation

Consider a dynamical system induced by the escort transformation from  $\mathbf{p}$  to  $\mathbf{P}$  defined by (5). When we identify the set of escort distributions  $\mathcal{E}^n$  with  $\mathcal{S}^n$ , the transformation is regarded to define a flow  $\mathbf{P}^{(t)}$  on  $\mathcal{S}^n$  parametrized by  $t \in \mathbf{R}$ :

$$P_i^{(t)} = \frac{(p_i)^t}{\sum_{j=1}^{n+1} (p_j)^t}, \quad i = 1, \dots, n+1, \quad \mathbf{P}^{(1)} = \mathbf{p} \in \mathcal{S}^n, \quad (14)$$

where  $\mathbf{p}$  is a *fixed* probability distribution.

Recalling the standard dually flat structure, which is obtained by limiting  $q \rightarrow 1$  (or  $\alpha \rightarrow 1$ ) in Proposition 1, we have the corresponding coordinates<sup>c</sup>  $\theta_{\mathbf{p}}^i := \theta^i(\mathbf{p}) = \ln(p_i) - \ln(p_{n+1})$ ,  $i = 1, \dots, n$ . In this case, if a curve  $(\theta^i(t))$  on  $\mathcal{S}^n$  is affinely parametrized by  $t \in \mathbf{R}$ , we call it *e-geodesic*.<sup>7</sup>

Since it follows that

$$\theta^i(t) := \theta^i(\mathbf{P}^{(t)}) = \ln P_i^{(t)} - \ln P_{n+1}^{(t)} = t(\ln p_i - \ln p_{n+1}) = t\theta_{\mathbf{p}}^i, \quad i = 1, \dots, n,$$

we conclude from a viewpoint of information geometry that the flow of the escort transformation (14) evolves along the e-geodesic curve that passes  $\mathbf{p}$  at  $t = 1$ .

Note that the arbitrary flows (14) converge to the uniform distribution independently of  $\mathbf{p}$ , when  $t \rightarrow 0$ . On the other hand, when  $t \rightarrow \pm\infty$ , it converges to a distribution on the boundary of  $\mathcal{S}^n$  depending on the maximum or minimum components of  $\mathbf{p}$ . See Ref. 41 as a relevant work. In several literature,<sup>42,43</sup> examples of physical models with a time-evolution of the power index of distribution functions are reported.

<sup>c</sup>These coordinates are called the *canonical parameters* in statistics literature.

The above result can be slightly generalized with a *projective transformation*  $\Pi_{\mathbf{r}} : \mathcal{S}^n \rightarrow \mathcal{S}^n$  defined by

$$\mathbf{p} = (p_i) \mapsto \Pi_{\mathbf{r}}(\mathbf{p}) := \left( \frac{r_i p_i}{\sum_{i=1}^{n+1} r_i p_i} \right), \quad i = 1, \dots, n+1,$$

for a given vector  $\mathbf{r} = (r_i) \in \mathbf{R}_+^{n+1}$ , and the relation with the replicator equation is elucidated.

**Proposition 4.** *For arbitrary  $\mathbf{r}$  the projective transformation of the escort flow given in (14) evolves along the  $e$ -geodesic curve that passes  $\tilde{\mathbf{r}} = \mathbf{r}/\|\mathbf{r}\|_1$  at  $t = 0$  and  $\Pi_{\mathbf{r}}(\mathbf{p})$  at  $t = 1$ . This flow evolves along the trajectory of the replicator equation (10) with constants  $f_i = \ln(p_i)$ ,  $i = 1, \dots, n+1$ .*

**Proof.** The first statement follows from direct calculation of coordinates  $\theta^i$  for the standard dually flat structure when  $q \rightarrow 1$  ( $\alpha \rightarrow 1$ ):

$$\theta^i(\Pi_{\mathbf{r}}(\mathbf{P}^{(t)})) = \ln(r_i P_i^{(t)}) - \ln(r_{n+1} P_{n+1}^{(t)}) = t\theta_{\mathbf{p}}^i + \ln(r_i/r_{n+1}), \quad i = 1, \dots, n.$$

To prove the second statement note that the flow  $\Pi_{\mathbf{r}}(\mathbf{P}^{(t)})$  is a normalization of a vector  $\mathbf{y}(t)$ , each component of which is  $y_i(t) = r_i(p_i)^t$ . Hence,  $\mathbf{y}(t)$  satisfies the following linear differential equation:

$$\dot{y}_i = \ln(p_i)y_i, \quad y_i(0) = r_i, \quad i = 1, \dots, n+1.$$

By setting  $x_i = y_i/\|\mathbf{y}\|_1$ , we have

$$\frac{d}{dt} \ln(x_i) = \ln(p_i) - \frac{1}{\|\mathbf{y}\|_1} \sum_{j=1}^{n+1} \dot{y}_j = \ln(p_i) - \sum_{j=1}^{n+1} x_j \ln(p_j), \quad i = 1, \dots, n+1.$$

Thus,  $\Pi_{\mathbf{r}}(\mathbf{P}^{(t)})$  is the solution of

$$\dot{x}_i = x_i \left( \ln(p_i) - \sum_{j=1}^{n+1} \ln(p_j)x_j \right), \quad x_i(0) = \frac{r_i}{\|\mathbf{r}\|_1}, \quad i = 1, \dots, n+1.$$

This proves the second statement.  $\square$

## 5. Concluding Remarks

We have discussed two applications of escort probabilities and the dually flat structure  $(h, \nabla, \nabla^*)$  on  $\mathcal{S}^n$  induced by conformal transformations of the  $\alpha$ -geometry. They are used to new directions except the studies of multifractal or nonextensive statistical physics.

We first demonstrate a direct application of the conformal flattening to computation of  $\alpha$ -Voronoi partitions and  $\alpha$ -centroids. Escort probabilities are found to work as a suitable coordinate system for the purpose. Further, conformal divergence and projective equivalence of affine connections also play important roles.

In behavioral analysis of dynamical systems we present the properties of gradient flows with respect to the conformal metric and discuss a relation with the replicator equation. Next, we show that the projective transformation of the escort flow is e-geodesic. This flow describes a time-evolution of the power index of distributions.

Physical interpretation of the obtained conformal structure is another future research direction.

## Acknowledgments

We would like to thank Prof. Tatsuaki Wada for helpful comments. A. Ohara was partially supported by JSPS Grant-in-Aid (C) 23540134 and AOARD Grant No. 10-1-4060.

## References

1. C. Beck and F. Schlögl, *Thermodynamics of Chaotic Systems* (Cambridge University Press, Cambridge, 1993).
2. C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World* (Springer, New York, 2009).
3. J. Naudts, *Generalized Thermostatistics* (Springer, London, 2011).
4. A. Ohara, *Phys. Lett. A* **370** (2007) 184.
5. A. Ohara, *Eur. Phys. J. B* **70** (2009) 15.
6. S.-I. Amari, *Differential-Geometrical Methods in Statistics* (Springer-Verlag, Berlin, 1985).
7. S.-I. Amari and H. Nagaoka, *Methods of Information Geometry* (AMS & Oxford, Rhode Island, 2000).
8. A. Ohara, H. Matsuzoe and S.-I. Amari, *J. Phys.: Conf. Ser.* **201** (2010) 012012.
9. S.-I. Amari and A. Ohara, *Entropy* **13** (2011) 1170.
10. H. Matsuzoe and A. Ohara, *Proc. 2nd Int. Colloquium on Differential Geometry and its Related Fields* (World Scientific, Singapore, 2011), p. 55.
11. K. Onishi and N. Takayama, *IEICE Trans. Fundamentals* **E79-A** (1996) 533.
12. K. Onishi and H. Imai, *Proc. 13th ACM-SIAM Symposium on Computational Geometry* (1997), p. 463.
13. J.-D. Boissonnat, F. Nielsen and N. Nock, *Discrete Comput. Geom.* **44** (2010) 281.
14. H. Imai and M. Inaba, *IEICE Trans. Inf. Syst.* **E83-D** (2000) 27.
15. G. Lebanon, *IEEE Trans. Pattern Anal. Mach. Intell.* **28** (2005) 497.
16. S.-I. Amari, *Neural Comput.* **19** (2007) 2780.
17. N. N. Čencov, *Statistical Decision Rules and Optimal Inference* (AMS, Rhode Island, 1982), [Originally published in Russian (Nauka, Moscow, 1972)].
18. H. Edelsbrunner and R. Seidel, *Discrete Comput. Geom.* **1** (1986) 25.
19. H. Edelsbrunner, *Algorithms in Combinatorial Geometry* (Springer-Verlag, 1987).
20. A. Fujiwara and S.-I. Amari, *Physica D* **80** (1995) 317.
21. S. Eguchi, *Ann. Stat.* **11** (1983) 793.
22. T. Kurose, *Math. Z.* **203** (1990) 115.
23. T. Kurose, *Tôhoku Math. J.* **46** (1994) 427.
24. T. Kurose, *Interdiscip. Inform. Sci.* **8** (2002) 89.
25. H. Matsuzoe, *Hiroshima Math. J.* **29** (1999) 175.
26. C. Tsallis, *J. Stat. Phys.* **52** (1988) 479.
27. A. Ohara, H. Matsuzoe and S.-I. Amari, arXiv:1010.4965v1 (2010).

28. P. T. Landsberg and V. Vedral, *Phys. Lett. A* **247** (1998) 211.
29. A. K. Rajagopal and S. Abe, *Phys. Rev. Lett.* **83** (1999) 1711.
30. T. Wada and A. M. Scarfone, *Phys. Lett. A* **335** (2005) 351.
31. K. Nomizu and T. Sasaki, *Affine Differential Geometry* (Cambridge University Press, Cambridge, 1993).
32. J. Naudts, *J. Inequal Pure Appl. Math.* **5**(4) (2004) 102.
33. A. Rényi, *Proc. 4th Berkeley Symp. Math. Statist. Probability*, Vol. 1 (University of California Press, Berkeley, 1961), p. 547.
34. H. Matsuzoe, in *Emerging Trends in Visual Computing*, ed. F. Nielsen (Springer, Berlin, 2009), p. 103.
35. F. Nielsen and R. Nock, *Int. Symp. Voronoi Diagrams (ISVD)*, DTU Lyngby, Denmark (IEEE Press, New York, 2009).
36. B. Chazelle, *Discrete Comput. Geom.* **10** (1993) 377.
37. J. Hofbauer and K. Sigmund, *The Theory of Evolution and Dynamical Systems: Mathematical Aspects of Selection* (Cambridge University Press, 1988).
38. S. Shahshahani, *A New Mathematical Framework for the Study of Linkage and Selection*, Memoirs of the AMS, Vol. 17, No. 211 (AMS, 1979).
39. N. Ay and I. Erb, *J. Dyn. Differ. Equ.* **17** (2005) 427.
40. M. Harper, *Physica D* **240** (2011) 1411.
41. S. Abe, *Phys. Rev. E* **68** (2003) 031101.
42. A. Taruya and M. Sakagami, *Phys. Rev. Lett.* **90** (2003) 181101.
43. A. Campa, P. Chavanis, A. Giansanti and G. Morelli, *Phys. Rev. E* **78** (2008) 040102(R).